

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

### STAFF USE ONLY

Date completed: 04-16-03  
Searcher: Beverly E 4997  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

#### Search Site

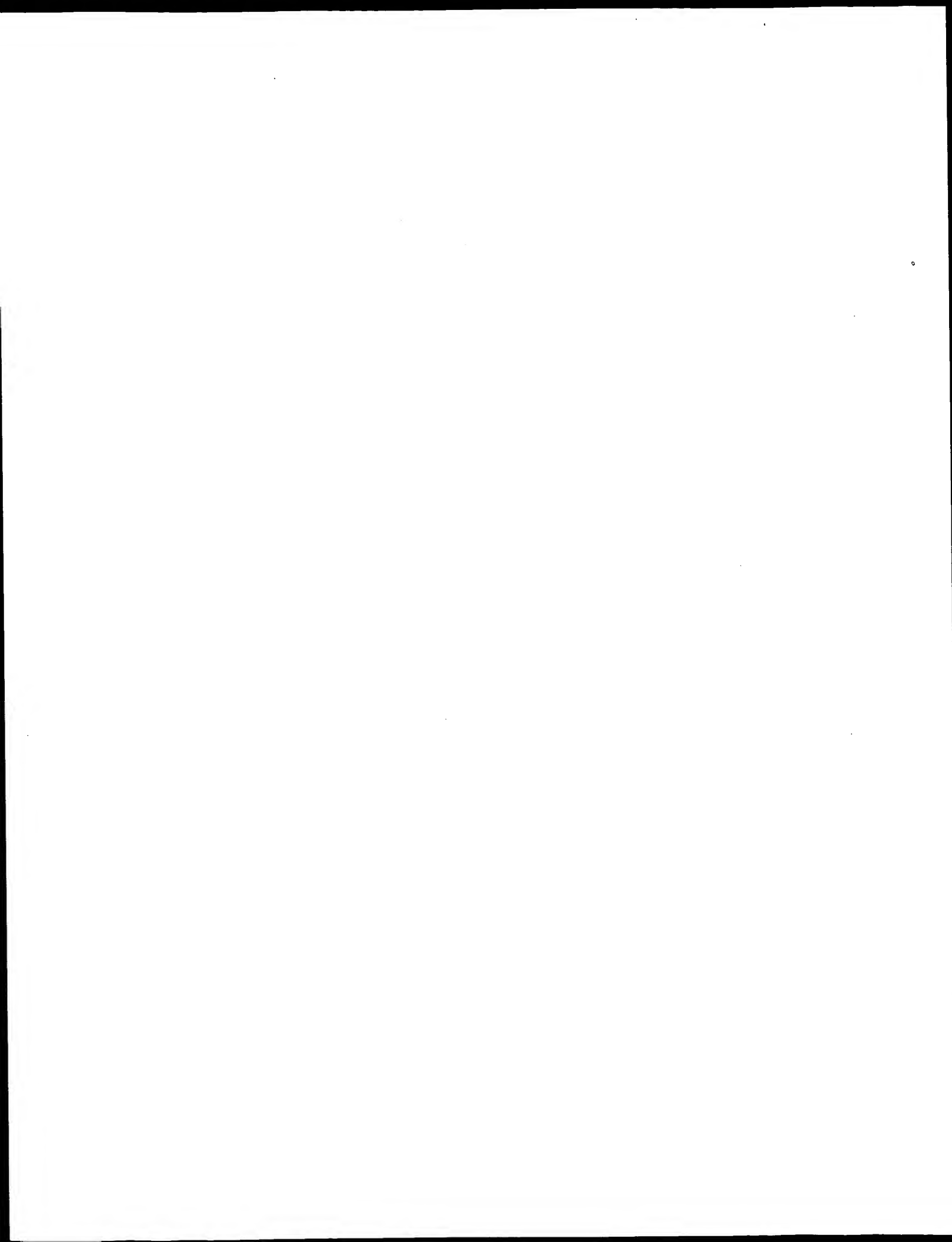
\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

#### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

#### Vendors

\_\_\_\_\_ IG Suite  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
\_\_\_\_\_ Other CGN



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:06:01 ; Search time 3378 Seconds

(without alignments)  
9175.396 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgatacccccagagatttaa.....cccttcaagctctctttaa 1065

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

GenEmbl:\*

- 1: gb\_da:\*
- 2: gb\_hlg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_vl:\*
- 14: gb\_vl:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hlg\_hum:\*
- 31: em\_hlg\_inv:\*
- 32: em\_hlg\_other:\*
- 33: em\_hlg\_mus:\*
- 34: em\_hlg\_pln:\*
- 35: em\_hlg\_rtd:\*
- 36: em\_hlg\_mam:\*
- 37: em\_hlg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hlgo\_hum:\*
- 40: em\_hlgo\_mus:\*
- 41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1065	100.0	1065	6	AR043346
2	1065	100.0	1065	6	AR078122
3	1063.4	99.8	12497	1	AE000771
4	145.4	13.7	12075	1	AE013160
5	103.8	9.7	12452	1	AE010265
6	90.8	8.5	11932	1	AE002508
7	90.8	8.5	349980	6	AX044033
8	89.2	8.4	20334	1	AE001764
9	87.6	8.2	326301	1	NMA622491
10	86	8.1	22014	1	AE000963
11	80.6	7.6	188050	1	AE646072
12	72.2	6.8	1480	1	AF246314
13	68.8	6.5	11265	1	AE012885
14	64.2	6.0	1285	8	AE011725
15	64.2	6.0	1570	8	AY050832
16	61	5.7	7218	6	16494
17	60.2	5.7	17492	1	U67539
18	60	5.6	1242	8	NPL278767
19	56.8	5.3	1556	1	NHRA
20	55.2	5.2	300600	1	AE005369
21	54.4	5.1	13190	1	AE013011
22	46.2	4.3	1085	6	AX431883
23	45	4.2	301150	1	AE003190
24	44.2	4.2	25770	1	AE000964
25	44.2	4.2	25770	1	AE000964
26	43.6	4.1	26335	1	CU11168X5
27	42.2	4.0	1092	6	AR078126
28	41.6	3.9	119184	1	D90909
29	40.8	3.8	10872	1	AE008031
30	40.8	3.8	11490	1	AE009065
31	40.6	3.8	189287	2	AE068362
32	40.6	3.8	200918	9	AC012360
33	40.2	3.8	177585	2	AC094167
34	40.2	3.8	226608	2	AC105612
35	40	3.8	65785	2	AC101401
36	40	3.8	91365	2	AP004076
37	40	3.8	148159	2	AL772390
38	40	3.8	158050	8	AF480496
39	39.8	3.7	147565	2	AC112868
40	39.8	3.7	176744	2	AC128581
41	39.8	3.7	201420	2	AC099110
42	39	3.7	11080	1	AE006160
43	39	3.7	215296	2	AC093469
44	38.8	3.6	12257	1	AE009793
45	38.8	3.6	177014	2	AC022269

## ALIGNMENTS

RESULT 1

LOCUS AR043346

DEFINITION Sequence 23 from patent US 5814473.

ACCESSION AR043346

VERSION AR043346.1 GI:5964354

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1065)

AUTHORS Warren, P. V. and Swanson, R. V.

TITLE Transaminases and aminotransferases

JOURNAL Patent: US 5814473-A 23 29-SEP-1998;

FEATURES Location/Qualifiers

AR043346 Sequence 23 from patent US 5814473.

AR043346 1065 bp DNA

AR043346 linear PAT 29-SEP-1998

```

source      1. .1065
            /organism="unknown"
BASE COUNT  344 a      220 c      252 g      249 t
ORIGIN

```

```
Query Match      100.0%; Score 1065; DB 6; Length 1065;
Best Local Similarity 100.0%; Pred. No. 1.4e-290;
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

OY	1	ATTATPACCCCGAGAGATTAAAGAACTTGAGCTGATCAAGACGGAGGCTACATCCCGCTCC	60
Db	1	ATGATACCCCGAGAGATTAAAGAACTTGAGAACTTGACAAAGCGAGGTCACTCCCGCTCC	60
OY	61	GTCAGGCTTTCCTCTAACGAATTCCTCCCTAGACATTTCCCGAGAGATTAANAACAAAGGCC	120
Db	61	GTCAGGCTTTCCTCTAACGAATTCCTCCCTAGACATTTCCCGAGAGATTAANAACAAAGGCC	120
OY	121	TTAGAGAATTTAAAAAGGTTCCCTTGACAAATACCACACCCTCGAAGGAAAGGTTA	180
Db	121	TTAGAGAATTTAAAAAGGTTCCCTTGACAAATACCACACCCTCGAAGGAAAGGTTA	180
OY	181	AAAGCGTTCTTGGGAGATTTTTCGGGGTTAAGGAAGAAATTTAGTCTCGGTAAAGCT	240
Db	181	AAAGCGTTCTTGGGAGATTTTTCGGGGTTAAGGAAGAAATTTAGTCTCGGTAAAGCT	240
OY	241	TCGAGCACTCATATATACCTCTCATAGAGCTTAGGTGAAGTTACATACCTCCGTTAC	300
Db	241	TCGAGCACTCATATATACCTCTCATAGAGCTTAGGTGAAGTTACATACCTCCGTTAC	300
OY	301	ATACCTGTTCACCTTTCCCATGTAGAGATTAAGTGGAAAGTCTCGGAAACCCCTC	360
Db	301	ATACCTGTTCACCTTTCCCATGTAGAGATTAAGTGGAAAGTCTCGGAAACCCCTC	360
OY	361	GTTAAGGTTTCAACTGAGCAAGAACTTTGATTTAAGTTAGAAAGATATTGAATTATA	420
Db	361	GTTAAGGTTTCAACTGAGCAAGAACTTTGATTTAAGTTAGAAAGATATTGAATTATA	420
OY	421	GAGAAAGAAAAACCGGTTCTCGGGTACTTGTGCTTACCCAAACACCCACGGGAAACCTC	480
* Db	421	GAGAAAGAAAAACCGGTTCTCGGGTACTTGTGCTTACCCAAACACCCACGGGAAACCTC	480
OY	481	TTTTTCCAGGGAAAGATTGAGAGATTAAGAAACGAGGGTGTTCCTGTGAATAGACGA	540
Db	481	TTTTTCCAGGGAAAGATTGAGAGATTAAGAAACGAGGGTGTTCCTGTGAATAGACGA	540
OY	541	GCCCTATCATTTACCTCCGAGAAACCTTTCTGGAACACCGGCTCAAAAAGGAGATACG	600
Db	541	GCCCTATCATTTACCTCCGAGAAACCTTTCTGGAACACCGGCTCAAAAAGGAGATACG	600
OY	601	GTAGTTTGTAGACACTTTCACAAATTCGATATGCGAGATTAAAGGTAGGGATTTTATA	660
Db	601	GTAGTTTGTAGACACTTTCACAAATTCGATATGCGAGATTAAAGGTAGGGATTTTATA	660
OY	661	GGGAAGGGGGAATGCTCGAATAATTAACAAGGTGAGACTCCCTTCAACGTGACTAC	720
Db	661	GGGAAGGGGGAATGCTCGAATAATTAACAAGGTGAGACTCCCTTCAACGTGACTAC	720
OY	721	CCCTCTCAGAGTATGCAAAAGTTTCCCTCAACGAGGGAAGAGATTTCTTAATGAAAG	780
Db	721	CCCTCTCAGAGTATGCAAAAGTTTCCCTCAACGAGGGAAGAGATTTCTTAATGAAAG	780
OY	781	ATACAGAGAGTTGTAACAGAGCGAGAAGAGATGTACGACGAATGACAAATTAAGAAAG	840
Db	781	ATACAGAGAGTTGTAACAGAGCGAGAAGAGATGTACGACGAATTAAGAAATTAAGAAAG	840
OY	841	GTTGAGGTTTTCGAGTAAGGCTTAATCTTGTGCTTTCAAAAGCCCTTACCCGCCAC	900
Db	841	GTTGAGGTTTTCGAGTAAGGCTTAATCTTGTGCTTTCAAAAGCCCTTACCCGCCAC	900
OY	901	GAGGTTATCAGAGCTACTGAAAAGGGATGCTCGACAGAAAGTATCTTACATGGA	960
Db	901	GAGGTTATCAGAGCTACTGAAAAGGGATGCTCGACAGAAAGTATCTTACATGGA	960

QY	961	GGAGCTCCAAAGATGCTCTCAGGGTAGCGTAGGGAACCGGAAGAAACACAAAGTTCTG	1020		
Db	961	GGAGCTCCAAAGATGCTCTCAGGGTAGCGTAGGGAACCGGAAGAAACACAAAGTTCTG	1020		
QY	1021	GAAGCAGCTCGAGAGATATAAATCCCTTTCACGCTCTCTTAA	1065		
Db	1021	GAAGCAGCTCGAGAGATATAAATCCCTTTCACGCTCTCTTAA	1065		
RESULT 2	AR078122	1065 bp	DNA	Linear	PAT 31-NUG-2000
LOCUS	AR078122	Sequence 23 from patent US 5962283.			
DEFINITION	AR078122				
ACCESSION	AR078122				
VERSION	AR078122.1	GI:10004868			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 1065)				
TITLE	Warren, P. V. and Swanson, R. V.				
JOURNAL	Transmisses and amphotransferases				
FEATURES	Patent: US 5962283-A 23 05-Oct-1999;				
source	Location/Qualifiers				
	1..1065				
	/organism="unknown"				
BASE COUNT	344 a	220 c	252 g	249 t	
ORIGIN					
Query Match	100.0%	Score 1065;	DB 6;	Length 1065;	
Best Local Similarity	100.0%;	Pred. No. 1,4e-280;			
Matches 1065;	Conservative	0;	Mismatches	0;	Indels
					Gaps
QY	1	ANGATACCCAGAGATTAAGGAAGCTTGAAGCTTACAAAGAGAGGTCACTCCCGCTCC	60		
Db	1	ATGATATCCCAAGAGATTAAGGAAGCTTGAAGCTTACAAAGAGAGGTCACTCCCGCTCC	60		
QY	61	GTGAGGCTTCTCTTACGAATTCCTTACGACTTCCCGAGAGATTAACAAAGGCGC	120		
Db	61	GTGAGGCTTCTCTTACGAATTCCTTACGACTTCCCGAGAGATTAACAAAGGCGC	120		
QY	121	TTAAGAAGATTAAAAAGGTCCCTTGAACAAATACCCAGACCCGAGAGGAGTAA	180		
Db	121	TTAAGAAGATTAAAAAGGTCCCTTGAACAAATACCCAGACCCGAGAGGAGTAA	180		
QY	181	AAAGCGTTCTTCGCGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTACGCT	240		
Db	181	AAAGCGTTCTTCGCGATTTTTCGCGGTTAAGGAAGAAATTTAGTCTCGGTACGCT	240		
QY	241	TCGAGACGAAGCTATATATACCTCTCAATATAGCATATAGGAACTTACATACCCGTTAC	300		
Db	241	TCGAGACGAAGCTATATATACCTCTCAATATAGCATATAGGAACTTACATACCCGTTAC	300		
QY	301	ATACCTGTTCCCACTTTTCCCATGTACAGATATAGTCCGAAGATTTCTCGGAAGACCCCTC	360		
Db	301	ATACCTGTTCCCACTTTTCCCATGTACAGATATAGTCCGAAGATTTCTCGGAAGACCCCTC	360		
QY	361	GTAAGAGTTCACTGACGAGAAACCTTGTATATAGACTTAGAAGAGATTTGAATTAATA	420		
Db	361	GTAAGAGTTCACTGACGAGAAACCTTGTATATAGACTTAGAAGAGATTTGAATTAATA	420		
QY	421	GAGAAAGAAACCCGTTCTCGGGTACTTTGTACCCAAACAAACCCACGCGGAAACCTC	480		
Db	421	GAGAAAGAAACCCGTTCTCGGGTACTTTGTACCCAAACAAACCCACGCGGAAACCTC	480		
QY	481	TTTTCAGGGGGAAGATGAGAGATTAAGAAACAGGGGTTTCTGTGTATATAGAGAA	540		
Db	481	TTTTCAGGGGGAAGATGAGAGATTAAGAAACAGGGGTTTCTGTGTATATAGAGAA	540		
QY	541	GCTACTATATATATATCCGAGAGAACTTCTGAGAGACGCGCTCAAAAGGAGATACG	600		
Db	541	GCTACTATATATATATCCGAGAGAACTTCTGAGAGACGCGCTCAAAAGGAGATACG	600		





/db\_xref="GI:2984294"  
 /translation="MKKHWALVSEGNIFQSLIKLVGGLITGSLISLIDGAIHSLSDA  
 TASLALFLSTKFEIKSERPPYGLYKLENGALVIAVLEFLFTAMEIIRQLKEGIN  
 FENLPIGIVTVLSLTLSELERAGKRLSPPLIADSYLTDADSSFLVLSL  
 SSYFGINIRRYAVAVALLIYVAFELKEQALIDISADKETVEIKRILISPE  
 VSEYKRLVNRAGRLFDIVATINTDDFKSHAIDEIERKIMKEIPNDWFIHVE  
 PCCKGTSTAVIAKODGVARFSDVKIIFPENEGNEPIETFKGDESCIAKELAR  
 RNDIVISGHTPIPSNRKAKWILKNNVWETKEENPEALSEVSQIKENT"  
 5148..6395  
 /gene="murD"  
 /note="aq\_2075"  
 5148..6395  
 /gene="murD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylmutamoylalanine-D-glutamate ligase"  
 /protein\_id="AAC07814.1"  
 /db\_xref="GI:2984290"  
 /translation="MKFLVGLGRSGKGLKLERGEFVYAGDDSQNPELMREVLGE  
 VDVVLSPGIPSPHLMKEALKEKEEYVGELEIAYREFKGVIALITGDKSTTRLT  
 YLILKPEPDEYFAGNIGKPESEVULENPGIAYLEVSFGKTLKFRNIGATISF  
 SVDLHDHPSIEDILKSKYRIFFENQTEDEFLINDIYVELTKKPSRARKYLSFIYN  
 SDVYFKDILFENKRLKIRGLHNVYASVSLALTLGLKPEDFEVYIEFPGILHR  
 LEFLGNNGVEVNDKSTPHALMALKTPEPDSVILIVGSKDKGADFEYSLRHIVK  
 KVIKIALIGETKEKIDSKWDITEVKTCTNLEAVKDLAREVSKLGNVLSFSPCSSPD  
 MFRNVEERGEKEFLVEIWFRT"  
 complement(6382..6996)  
 /gene="lrpP"  
 /note="aq\_2076"  
 complement(6382..6996)  
 /gene="lrpP"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoribosyl anthranilate isomerase"  
 /protein\_id="AAC07811.1"  
 /db\_xref="GI:2984287"  
 /translation="MVKVKIGITNLEDAFCAGADYICVITYPKSPRYTSKREL  
 EIIKALELNVKVAVVNVEPYEFIKELDGFLLDIGHDEDELKRGVLDGVIVY  
 FRVKEIPKVEGEMEKAAVILDTFSKEAYGSGTGFPMWEIARKLVEEGHKVFLSGGLN  
 PENYKEGTEFEPYGVIVSSGVEKEKKDKKVRREFIKTKSE"  
 7015..8556  
 /gene="snf"  
 /note="aq\_2077"  
 7015..8556  
 /gene="snf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Na(+):neurotransmitter symporter (Snf family)"  
 /protein\_id="AAC07817.1"  
 /db\_xref="GI:2984293"  
 /translation="MEVREHMAIRFLGILAMAGNAVGLNFLRPYQAAENGGAEM  
 IPYIATFLVGLPIEMWAMRGYAGAGHGTTPAIEYILRNRPFAKILVFGMLPI  
 VYAIYVYVESMTGFAIKFLVGLPEPPNATDPSILRPFKEIKSYIGVPGDEP  
 ILKPSLEAYIVFLITMPTINVELLIRGISKGIERRAKIAMPFLILAVTVRVLLIET  
 PNGTADGLNPLTPDEFELKDPGVIAVAGQIFPLISLGGALITTAASYRKQDIY  
 LSGITAAULNEKAEVILGGISIPAAVAFVANAVALAKAGANLGLTLPALFSOT  
 AGGTGLGLMEFLFAGLTSSIAIMQPIAFLEDELKLSKHAVALMTAAIVFSAHL  
 VWEFLKSLDEMDENAGTIGVVFGLTELIFEFWILFGADKAMEEINRGCIKVPYIY  
 VMRYTPAFVAVLAVMARVYIPKIMETHTMTWITTRYIIGLEFLFLFIVELAERRR  
 NHESA"  
 8543..9394  
 /gene="sppA"  
 /note="aq\_2080"  
 8543..9394  
 /gene="sppA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="proteinase IV"  
 /protein\_id="AAC07815.1"  
 /db\_xref="GI:2984291"  
 /translation="MRVLKMLILFSLILFSVKEFLISKFPVGHRIAVLKLGSVI  
 NPPLVNTKERLKDKNVAVIAYLRVDSGSGVAGASQEIYRAIERKFSKGVVISMG  
 VASGGIYSAAGNIIFANEGITTSIGVILIOHVAKKLVEKLGIEAIAITGKFKDT

LSPFRELTOEKEYLQNLVNEALQFIQAILIKYRSNKISEEKLREIADGRIFTGROAL  
 ELGIVDELGSIVDAIEKAKELSGYDAPVPFIOEERKSLKRLVADLEIEIKIINAYTP  
 LFLYINDY"  
 9672..11285  
 /gene="aq\_2082"  
 9672..11285  
 /gene="aq\_2082"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC07820.1"  
 /db\_xref="GI:2984296"  
 /translation="MKRVLEFGFGFGLVFGFVFEVMSILGLHLEHPPLYNAGF  
 ASVFLVLSFLNLINWVKALIKDYKKGVSVAFYDFLASIKLAIETMLVGLISML  
 GSTYIKONSFEMVLDQGYDGIWIMKMLMDVDFHSWYILFVILAVNLIPCSIKR  
 LPRVMQAFSKREIRILKLEHAKHLKPLTVKIPDKVKLEFLKKGFEVFEENKRL

Query Match 99.8%; Score 1063.4; DB 1; Length 12497;  
 Best Local Similarity 99.9%; Pred. No. 5.1e-290;  
 Matches 1064; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGATACCCAGAGATTAAAGACTTGAAGCTTAAAGACGAGAGGTCACTCCGCCCTCC	60
Db	12335	ATGATACCCAGAGATTAAAGACTTGAAGCTTAAAGACGAGAGGTCACTCCGCCCTCC	12276
QY	61	GTCAGGCTTCTCTAAGCAATTCCTACGACTTCCCGAGAGATTAACAAGAGGCC	120
Db	12275	GTCAGGCTTCTCTAAGCAATTCCTACGACTTCCCGAGAGATTAACAAGAGGCC	12216
QY	121	TTGAACAATTAATAAAGGTTCCTTGAACAATTAACCCAGACCCGAGAGAGATTA	180
Db	12215	TTGAACAATTAATAAAGGTTCCTTGAACAATTAACCCAGACCCGAGAGAGATTA	12156
QY	181	AAAGCGTTCCTTGGGATTTTTCGGCGTTAAGAGAAATAATTAGTCTCGTAAACGT	240
Db	12155	AAAGCGTTCCTTGGGATTTTTCGGCGTTAAGAGAAATAATTAGTCTCGTAAACGT	12096
QY	241	TCGAGCACTCATATCTACCTCTCATATAGCTATAGTAACTTACATACCGTTTAC	300
Db	12095	TCGAGCACTCATATCTACCTCTCATATAGCTATAGTAACTTACATACCGTTTAC	12036
QY	301	ATACCTGTCCACCTTCCCATGATAGATTAAGTSCGAAGTTCTCGGAAGACCCCTC	360
Db	12035	ATACCTGTCCACCTTCCCATGATAGATTAAGTSCGAAGTTCTCGGAAGACCCCTC	11976
QY	361	GTAAGGTCAACTGACGAATACTTGATATAGACTTGAAGAAGATTTGATTAATA	420
Db	11975	GTAAGGTCAACTGACGAATACTTGATATAGACTTGAAGAAGATTTGATTAATA	11916
QY	421	GAGAAAGAAACCCGTTTCGGGTACTTGTCCCAACCAACCCACGGAACCTTC	480
Db	11915	GAGAAAGAAACCCGTTTCGGGTACTTGTCCCAACCAACCCACGGAACCTTC	11856
QY	481	TTTTCCAGGGGAAGATTGAAGATTAAGAAACAGSGGTGTTTCTGTGTAATAGAA	540
Db	11855	TTTTCCAGGGGAAGATTGAAGATTAAGAAACAGSGGTGTTTCTGTGTAATAGAA	11796
QY	541	GCTACTATCACTTCTCGGAGAAACCTTCTGGAAGACGCGCTCAAAAGGAGATACG	600
Db	11795	GCTACTATCACTTCTCGGAGAAACCTTCTGGAAGACGCGCTCAAAAGGAGATACG	11736
QY	601	GTAGTTTGGAGACACTTCAAAATCGGATAGCGGATTAAGAGGATTTTATATA	660
Db	11735	GTAGTTTGGAGACACTTCAAAATCGGATAGCGGATTAAGAGGATTTTATATA	11676
QY	661	GGGAAGGGGAATCGTCTCAGAAATTACAGGTGAGACTCCCTTCAACGTGACCTAC	720
Db	11675	GGGAAGGGGAATCGTCTCAGAAATTACAGGTGAGACTCCCTTCAACGTGACCTAC	11616
QY	721	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAGGAGGAGAGAGATTCATATGGAAG	780
Db	11615	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAGGAGGAGAGAGATTCATATGGAAG	11556

ethanolicsus], score 237, E-value 8.00E-62"  
/codon\_start=1  
/transl\_table=11  
/product="Phosphoribosyl-AMP cyclohydrolase"  
/protein\_id="AAW25297.1"  
/db\_xref="GI:20517145"  
/translation="MIDIKELKFDEKGLIPAVDDYTRQVLMAYMNESSLKLTLEK  
GEYFFPSRSGIKMHGERTSGNTQVKRIEYDCDEADALLIOVEAKSPAGCHTNSCY  
RSNDETEDEGIELINKIYERIKGRKINVEBSYNNYLFPEKIDKILKKVGEPEATEYVI  
ASKNDKSDIIEVSDLIYHLMVLYENKNTLNDIYNLEKRYAK"  
complement(151.417)  
/gene="HisP"  
/note="Pfam match to entry PRA-PH, Phosphoribosyl-ATP  
pyrophosphohydrolase, score 159.1, E-value 7.50E-44"  
complement(445.678)  
/gene="HisI"  
/note="Pfam match to entry PRA-CH, Phosphoribosyl-AMP  
cyclohydrolase, score 186.6, E-value 4.10E-52"  
complement(777.781)  
/gene="HisI"  
/gene="HisP"  
complement(780.1557)  
/gene="HisP"  
/note="TME2133"  
complement(780.1541)  
/gene="HisP"  
/note="Best Blastp hit =  
g1|6175608|gb|AAFP05094.1|AFI50930\_3 (AFI50930) cyclase  
[Thermoanaerobacter ethanolicsus], score 365, E-value  
1.00E-100"  
/codon\_start=1  
/transl\_table=11  
/product="Imidazoleglycerol-phosphate synthase"  
/protein\_id="AAW25298.1"  
/db\_xref="GI:20517146"  
/translation="MLAKRIIPCLDVKDGRVKGINPVLKDAQDPVETAEERNELGA  
DELVEVLTASIEKKRIIMDIVKRRSEKIPPLTVGGISTDIDIEYELAKGADKRSI  
NMQAKOPKTLRQALPRGSCVYVAIDAKKRPDGTNYINGRINFGDLAENAK  
KKKIDGAEIITLSMDKDGITKDGYDIELTRLSIAVSIPVISAAGAGKPEHREYFTQ  
GAADALAAVPHVETDILKRLRYLKDEGIPVRL"  
complement(810.1532)  
/gene="HisP"  
/note="Pfam match to entry His biosynth, Histidine  
biosynthesis protein, score 458.7, E-value 4.40E-146"  
complement(1542.2279)  
/gene="HisA"  
/note="TTE2134"  
complement(1542.2264)  
/gene="HisA"  
/gene="HisP"  
/note="Best Blastp hit =  
g1|6175607|gb|AAFP05093.1|AFI50930\_2 (AFI50930)  
Phosphoribosyl formimino-5-aminomidazole isomerase  
[Thermoanaerobacter ethanolicsus], score 186, E-value  
3.00E-46"  
/codon\_start=1  
/transl\_table=11  
/product="Phosphoribosylformimino-5-aminomidazole  
carboxamide ribonucleotide (ProFAR) isomerase"  
/protein\_id="AAW25299.1"  
/db\_xref="GI:20517147"  
/translation="MLVTPAIDILIDKCVRLTRGNDSKEVYTYNDPAMAKKMWDEGGA  
KRHHVDDGKRQGHLYNKRVTIEKIVNSCVDDIEVGGIRNNEALDYLFSTGVSTIL  
GSAAYIDDDLIVSVSHYGRKTVIGIDSKRREVAWGMEERIKDTELAEAKIKGIGI  
KTIIFTDSDKGLHGPNPALKMDKMGVEVIAISGSIISIEDLKLKMGVGTAAIG  
KALYTGIMDLAAALLEEREGI"  
complement(1553.1557)  
/gene="HisP"  
complement(1550.2264)  
/gene="HisA"  
/note="Pfam match to entry His\_biosynth, Histidine  
biosynthesis protein, score 219, E-value 7.20E-62"  
complement(2255.2878)  
/gene="HisP"  
/note="TME2135"

```

CDS      complement(2255. .2866)
        /gene="Hish"
        /note="Best Blastp hit = g11743913863|sp|O66943|HIS5_AQUAE
AMIDOTRANSFERASE HISH g117433986|p|I|C70364
amidotransferase Hish - Aquifex aeolicus
g11298331|gb|AAC06901.1| (AE000705) amidotransferase Hish
[Aquifex aeolicus], score 207, E-value 7.00E-53"
/codon_start=1
/transl_table=11
/product="Glutamine amidotransferase"
/protein_id="AAM25300.1"
/db_xref="GI:20517149"
/translation="MIGIDYGMGNLRSVEKALKYGVDAKITSDAKEIKNFALLP
GVGAPPAIRVLITGMAREIKIHIEKIGPGLGICMOLFEKSEFEGEGLIKPK
GNVGLPKGNKIPHYGNNSLTIKIDTPKIGKGDYVPHSYVNGNKDEYAVAT
EYGEIPAVIITKDVFCQFPEKSGETGLKILNFGEMIKCL"
complement(2267. .2860)
/gene="Hish"
/note="Pfam match to entry GATase, Glutamine
amidotransferase class-I, score 176.7, E-value 3.80E-49"
complement(2275. .2279)
/gene="Hish"
complement(2874. .2878)
/gene="Hish"
complement(2885. .3482)
/gene="Hish"
/note="TTE2136"
complement(2885. .3469)
/gene="Hish"
/note="Best Blastp hit = g117387773|sp|O33564|HIS7_RHOSH
IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGD)
g112398791|emb|CA60712.1| (X87256)
imidazoleglycerol-phosphate dehydratase [Rhodobacter
sphaeroides], score 223, E-value 2.00E-57"
/codon_start=1
/transl_table=11
/product="Imidazoleglycerol-phosphate dehydratase"
/protein_id="AAM25301.1"
/db_xref="GI:20517149"
/translation="MRKAVERKRTAETDIYELINDGSDYDIKIGFEDMLSLPA
KHGLPDKIVIAKGLEDVTHHVEDVGLGATFKAGDSKIRFETFPVPEAL
VRVSLDISGRPLIYDPLKAEKRGNEPTEVEEPRFAVFPGLTLHVELLHGVNH
HIEASFALGKADDEALKDERIDIGISTGIL"
complement(2945. .3379)
/gene="Hish"
/note="Pfam match to entry IGD,
imidazoleglycerol-phosphate dehydratase, score 293.4,
E-value 6.20E-100"
complement(3466. .4534)
/gene="Hisc2"
/note="TTE2137"
complement(3466. .4521)
/gene="Hisc2"
/note="Best Blastp hit = g117450424|p|I|E70478
histidinol-phosphate aminotransferase - Aquifex aeolicus
g112984289|gb|AAC07813.1| (AE000771) histidinol-phosphate
aminotransferase [Aquifex aeolicus], score 230, E-value
2.00E-59"
/codon_start=1
/transl_table=11
/product="Histidinol-phosphate aminotransferase/Tyrosine
aminotransferase"
/protein_id="AAM25302.1"
/db_xref="GI:20517150"
/translation="MIENLRREIKGFKNVEVNYPYKKMDANETPEELPEEVKNI
GTVIVSIHNYITPDPAEKLEBELARCSVYPKNIFVNGSDEITHLIMLAFVKGDT
VLYPHEFSAMSIYISKIAGANETAVNLNEDYTYNVEREAVNERKPKLVPCNPNP
TGSVIDEEDIRIRIEKARGIVIVDAVEFEYGLTVLPYIDRENLIVRTSKAGTIA
KLGVGVALSNGEIVLYLNLKSPYNLNSLSQRIAEVLKSVGLKERVNTLINERKLV
GLKANKINGIKVPSHANEVLCFENANDVHKRLVERGILVRNFSVKGIEGLRTIVS
SSDANDYLINALREILS"
complement(3469. .4515)
/gene="Hisc2"

misc_feature
        /note="Pfam match to entry aminotran_1, Aminotransferase
class-I, score -25.9, E-value 1.60E-08"
        complement(3478. .3482)
        /gene="Hish"

RBS      Query Match      13.7%; Score 145.4; DB 1; Length 12075;
        Best Local Similarity 49.8%; Pred. No. 7.5e-30;
        Matches 511; Conservative 0; Mismatches 501; Indels 15; Gaps 5;

        /note="Pfam match to entry aminotran_1, Aminotransferase
class-I, score -25.9, E-value 1.60E-08"
        complement(3478. .3482)
        /gene="Hish"

16  ATTAAGAACTTAAGCTTACAGAGGAGGTCACGCCCCCGCTCCAGCTTCTCT 75
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4494  ATTAAGAACTTAAGCTTACAGAGGAGGTCACGCCCCCGCTCCAGCTTCTCT 75
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

76  AACGAATTCCTTACAGCTTCCCGAGAGATTAACAAAGGCGCTTGAAGATTAA 135
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4434  AACGAGACCCCTTTTGAACCTTCCGAAAGAGTAATGAAAATATAGGGATTGTAAG 4375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

136  AAGGTTCCCTTGAACAAATACCCAGACCCCGAAGCCGAAGAGTTAAAGCGTTCTGG 195
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4374  TCTATCCATGTAAATATATATCCAGACCCCGAAGAAATTTGAGAGAGTTAGCG 4315
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

196  GATTTTTCGCGCTTAAAGAAATTTAGTTCGTAACGTTCCGAGCACTCAT 255
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4314  AGATATTGCACTGCTCACTCTAAATAATTTTGTAGAAATAGTTCTGATGAATATA 4255
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

256  TACTACCTTCATATAGCTATAGTGAACCTTACATACCGCTTACATACCTGTTCCAC 315
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4254  CACCTTATATAGCTTGGCTTGTGATAAAGGGATACAGTTTGTACCTTCCCATCT 4195
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

316  TTTCCCATGTACGATTAAGTGAAGGTCGTAAGAGACCCCTCGTAAGGTTCAACG 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4194  TTTCCCATGTACGATTAAGTGAAGGTCGTAAGAGACCCCTCGTAAGGTTCAACG 4135
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

376  GACCAAACTTGTATATAGCTTATAGAAAGATTTGAATTAATAGAGAAAGAAACC 435
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4134  AATAGAGCATTAATTAATATGTTGAAGATTTCCAGACAGTGAAGATTAAGGCT 4075
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

436  GTTCTCGGACTTGTGCTTACCCCAACACCCCGAAGACCTCTT--TCAGGGGA 493
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4074  AAATTTGGTTTGTGCAATCAACAAATCCCTGGAAGCGTAAATAGTGAAGAGAT 4015
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

494  AGATTGAGAGATTAAGAAACAGG-GTGTCTTCTGTTATATAGCAAGCTCTATCAT 552
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4014  AATATGCAATTAATTTGAAGAAAGCAAGAGGATTTATATGATGATGAAGCTTATTGAA 3955
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

553  TACTCCGAGAAACCTTCTGGAAGACGCGCTCAAAAGGAGATACGGATGTTTACG 612
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3954  TTTATGGAAGAAACCTTGTGCGCATATATAGATGATTTGAAATTTGATTTGCTGCGA 3895
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

613  ACACCTTCAAAA--ATCGGTATGCGAGTTAAGGATGAGATTTAATAGGAAGGCG 669
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3894  ACATTGTCAAAAGCTTTGGGATTGCAAGATTAGAGTGGGATATGCTCTCTCAATGCA 3835
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

670  GAAATCGCTCAAGAAATTAACAGAGTGAGACTCCCTTCAACGTGACCTACCCCTCAG 729
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3834  GAAATGAGAAATTTTAATCTTGTAAATCTCCTTAATATCACTCTTATACAG 3775
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

730  GTGATGCAAAAGTCTCTCAGGAGGAGAAAGATTTCCATAGGAAAGATACAGAG 789
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3774  AGAATGACATTAAGATTTTAAA-----AAGTGTGTTTAAAGGAGAGGATTAATTAC 3721
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

790  GTTGAACAGAGCCGAGAAAGATGACAGAAATGAAGAAATAGAGAGATGAGAGTT 849
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3720  ATTTAAACGAGAGGAAAGAAAGCTGTAAGAGATTAATTAATTAATTAATGATGATTAAGTG 3661
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

850  TTTCCAGTAAGGCTTAACCTTGTCTTTCAGAACGCTTACCCGCCACAGAGTTTAT 909
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3660  TATCTTCTCATGCAATTTTGTCTTT--GTAATTTGAAGATGCAATGATGATGACAC 3604
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

910  CAGAGCTACTGAAAGAGATGCTCTCTGACGAAGATCTTACATGGAAGAGCTCAA 969
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3603  AAAAGACTTGTGAAAGGGAATTTTGGTGAAGAAATTTTCTAATGTGAAGGGGTTGGA 3544
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

gene  
DSD  
gene  
DSD  
gene  
DSD  
gene  
DSD  
gene  
DSD

Histidine family: (hisG)"  
/codon\_start=1  
/transl\_table=11  
/product="ATP phosphoribosyltransferase"  
/protein\_id="AA81782.1"  
/db\_xref="GI:18893815"  
/translation="MRFTYPRKRLYQDSIKVLEAGIKYRKDERALIMGDNNEFLIL  
ARAFDVVEHGDIDIGASDVVEROSDFIPIELPEEKCRLSTAVREKVDPEN  
MDGKIATKTPNTKYFENLNIYEVIKLHGSLTAPRIGADALVDIVETGNLR  
NGYLEAKIMDYNSALLNRIISQKKFEFEINSITFKIRRSVGH"  
1614..2741  
/gene="pf1659"  
1614..2741  
/gene="pf1659"  
/EC\_number="3.5.4.19"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family"  
/codon\_start=1  
/transl\_table=11  
/product="histidinol dehydrogenase"  
/protein\_id="AA81783.1"  
/db\_xref="GI:18893816"  
/translation="MDIEYVEELLKDIRERGLEAIRYSKKFDGECEPFESKEEF  
EAERIPERKETIKRLTRIEKYHRQLKODILIENGSLYGLYKPLKRIGIYVP  
GRLPSLTLMWGPVARVAQVEIVTVNPCKGKNAPIITYIAKLKIEFYVLGGIO  
IAAAMAGICKMKOKJFGCNKFVNKAEROVFQVIGDSIAGSEAYIADSAAKDE  
LVLDLSOLEHGSDKAMLLITTSKEDIQCRBSGITRYNSSICECCAKNPATPHH  
EIITRRPKLIEDINACGIYLGYTPPADYDTGNHHLPFGCAKRMGVLYTWDE  
MKPTSIAMSREFPITSERBLGRIELEBTELRKKSLEVAK"  
2808..3338  
/gene="pf1660"  
2808..3338  
/gene="pf1660"  
/EC\_number="4.2.1.19"  
/note="Function Code: 1.7 Amino acid Biosynthesis:  
Histidine family"  
/codon\_start=1  
/transl\_table=11  
/product="imidazoleglycerol-phosphate dehydratase"  
/protein\_id="AA81784.1"  
/db\_xref="GI:18893817"  
/translation="MRRRTKEDTLIVLEGKGKITNDLTIHMHTAFAFYLGDMR  
TAHYDRHHIMEDIGITIGEALRENLEPKPTRGNMAIMPDDALVYSDISNRITA  
VDWNIDAEGRVASULKEFWGLARGLRATIHTKOLSGNAHHIVEAFAKGLGWALR  
VAKRESERVESTRGVL"  
3338..3922  
/gene="pf1661"  
3338..3922  
/gene="pf1661"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family (hisH)"  
/codon\_start=1  
/transl\_table=11  
/product="glutamine amidotransferase"  
/protein\_id="AA81785.1"  
/db\_xref="GI:18893818"  
/translation="MDRIAIVDLGICLANVKALKGIYTSPDIETEKADKTIVLPQG  
NSEAVVPMGLAKDKDITIEGINEGRPUGICLGMQLPEESESSPGEGIGTFKGVMW  
LKWRVRPDHGAQNMVIRKECKLFGKLNGSYFFVSHYAVPODDPIIATTDVENME  
FVSVCVENIFSVQHPERSKKNGLILRNFRRL"  
3924..4646  
/gene="pf1662"  
3924..4646  
/gene="pf1662"  
/gene="pf1662"  
/EC\_number="3.5.4.19"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family"  
/codon\_start=1  
/transl\_table=11  
/product="n-(5'-phospho-d-ribose)ylformiminol-5-amino-1-  
(5'-phosphoribosyl)-4-imidazolecarboxamide isomerase"  
/protein\_id="AA81786.1"  
/db\_xref="GI:18893819"  
/translation="MKFRIPYPAIDIMDGCVVRLYKGGKDEQVYKVGDPULKFEKSESEVV

DKIHIVDLDGAFSGKPNLDIVKIPDVGVVOIGGGEDYSEIKKAYDIGENVII  
GTRKADISTLERITSEFEGITVSLDSKGIETTKMEIEEELVDAKMKRKPNRP  
VTVIDROGTELEIKIRKFMGDEEPIYAGVATPTDITLOAEIGFSVITGKALYEC  
TISLEALEVAAVSGKNVSSS"

gene  
4570. .5370  
/gene="pf1663"  
4570. .5370  
/gene="pf1663"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family; (husef)"  
/codon\_start=1  
/transl\_table=11  
/product="imidazoleglycerol-phosphate synthase, cyclase  
subunit"  
/protein\_id="AA181787.1"  
/db\_xref="GI:18893820"  
/translation="MKELSLKKMLRLKVLARRIIAALDITKGRVYGVGKIRKIRDA  
GPIELAKRYESEGIDIVFDITVASYERKLLDLVERVAEIVPTVGGIITVE  
EAREILIRGADKIFINTAVENPSLIKELADIGSANVLAIDAMNKRKFEVYTHG  
RKPRGDVAEMAKRVDELGAGEILITSMOTDGTGKGPDIPLTKAVADVNIPIVIA  
AGGAKPEHLEFVIGSDALMAASIFHYGEYVRELKMLAKNGIPVRLD"

gene  
5345. .6010  
/gene="pf1664"  
5345. .6010  
/gene="pf1664"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosyl-AMP cyclohydrolase /  
phosphoribosyl-ATP  
protein\_id="AA181788.1"  
/db\_xref="GI:18893821"  
/translation="MEFOLDLRGEHEMETEKLLQVNMKNNGIPIVIVODEGEVLT  
LAYMKKEARLKELEIVAHYSGRIGRIGKGEVNIQTVKIEIDDDNALLITS  
PKPRGHTNYSCTFYRKLEGEPRVLPYKSSILKELEIRKRETEVGGSTKLP  
KSGREKTKKFGEEALEVLAERREKITEVADLYHLVLITVDISGEVMEILRR  
RRK"

gene  
6007. .7023  
/gene="pf1665"  
6007. .7023  
/gene="pf1665"  
/note="Function Code: 1.7 Amino Acid Biosynthesis:  
Histidine family"  
/codon\_start=1  
/transl\_table=11  
/product="putative histidinol-phosphate aminotransferase  
(imidazole)"  
/protein\_id="AA181789.1"  
/db\_xref="GI:18893822"  
/translation="MIMEILNPEPYRAVEGNRYIMWDKNESPVDLPQKLEELLEEL  
KRIEFNRYPHITSDPLRELALEYGDKENINAVNGSDLIYLVKMRGKXIVVTS  
TQGMYSFPAKLGIPYKIDPLKEDFTIDEREIAEGKASAFISPNPNQSPEN  
NEKIVLDSGRVILDEAVSEFSKSPITISEENLVITRTSEKAGIAGICGYMA  
NEKIIDALYKRIILPYNLNLMVAIKMEHDIYKRIKILIKERERREFFITYS  
PSRANFLMKLDIYDLTKKGIYVKKLSGRLSGHRTVITGKRWENDELKALKEFLE  
CRG"

gene  
7014. .7730  
/gene="pf1666"  
7014. .7730  
/gene="pf1666"  
/note="Function Code: 16.1 Conserved Hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"

Query Match 9.7%; Score 103.8; DB 1; Length 12452;  
Best Local Similarity 51.9%; Pred. No. 4.7e-18;  
Matches 290; Conservative 0; Mismatches 257; Indels 12; Gaps 2;

65 GCGTTTCCTCAACGATTCCTCCGAGAGATAAACAAGAGGCGCTTAG 124  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6071 GCGTTGATAAATGAAGTCCCTATGACCTCCCTCCCACTTAAGAGAGATATTAG 6130

QY 125 AAGATTAAAGAGTCCCTTGAACAATATCCAGACCCCGAAGAGATTAAAG 184  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6131 AAGACTAAGAGATATAGATTATATCTCATATACCTAGACCCCTTAAGAG 6190  
QY 185 CGGTTCTTGGGATTTTTCGGCGTTAAGAGAATAATTATGTTCTGTAACGTTGG 244  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6191 AGGCCCTGCAGAAATTTACGAGACTAAAGAAACATAGCCGTTGGAAATGTAGTG 6250  
QY 245 ACGAATCTATATCTACTCTCATAGCTATAGTGAACCTTACATCCGTTACATAC 304  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6251 ACGAATCTATATATTTATCTGTGAAGATGTCAGAGGAATCATATGTGTT-----A 6304  
QY 305 CTGTCCCAACCTTCCATCTAGCAGATAGTGAAGAAAGTTCTGGAAGACCCCTGTAA 364  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6305 CTTCCCAACCTTTCGATGATATCATTTTTCGGAAGCTTCATGAAATTCAGTGAAG 6364  
QY 365 AGGTTCAACCTGAGCAAACTTTGATATAGCTTAGAAGAAGTATTAATTAAGA 424  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6365 ACATTCCCTTAAGAGGAGCTTCAACATTTGATGAGAGAGAAATGACAGAAAGAAAG 6424  
QY 425 AAGAAACCCGTTCTCGGTACTTGTCTTACCAACACCCGAGGAAACCTTTT 484  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6425 CAGCATCAGCAATT-----TTCATAGCTTCAACCAACACCCGAGGAAACCTTCAAC 6478  
QY 485 CCAGGGAAGATGAGAGATAGAAACAGGGGTGTTTCTGTGTATATAGACAGCCT 544  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6479 CAGAAATGAAGTCTCAAGATTCATGACCTCGGAGAGATGATATCTGATGAACCT 6538  
QY 545 ACTATCATCTACCCGCAAAACCTTTCGGAAGACCGGCTCAAAAGGAATATCGTAG 604  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6539 ATTCGAGCTTTCGGGAAAGCTTCATCAACCAAAATCAGTAGATGAATTAAGTAA 6598  
QY 605 TTTTGAGACACTTTCACAA 623  
11111 11111 11111 11111 11111 11111 11111 11111  
Db 6599 TTCTAAGAGCGTTTCAAA 6617

RESULT 6  
AE002508 11932 bp DNA linear BCT 25-MAY-2000  
LOCUS  
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 150 of 206  
of the complete genome.  
ACCESSION AE002508 AE002098  
VERSION  
KEYWORDS AE002508.1 GI:7226820  
SOURCE  
ORGANISM Neisseria meningitidis MC58.  
Neisseria meningitidis MC58.  
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
Neisseria.  
REFERENCE  
AUTHORS 1 (bases 1 to 11932)  
Tetzelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
Dougherty,B.A., Mason,T., Parksey,D.S., Blair,E.,  
Citron,H., Clark,E.B., Cotton,M.D., Uitterback,T.R., Khouri,H.,  
Oin,H., Vamathevan,J., Gill,J., Scarlato,V., Massignani,V.,  
Piazza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
Rappuoli,R. and Venter,J.C.  
Complete genome sequence of Neisseria meningitidis serogroup B  
strain MC58  
JOURNAL Science 287 (5459), 1809-1815 (2000)  
MEDLINE 20175755  
PUBMED 10710307  
REFERENCE 2 (bases 1 to 11932)  
Tetzelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
Dougherty,B.A., Mason,T., Parksey,D.S., Blair,E.,  
Citron,H., Clark,E.B., Cotton,M.D., Uitterback,T.R., Khouri,H.,  
Oin,H., Vamathevan,J., Gill,J., Scarlato,V., Massignani,V.,  
Piazza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
Rappuoli,R. and Venter,J.C.

Piazza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,  
 Rappuoli, R. and Venter, J.C.  
 Direct Submission  
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

## FEATURES

Location/Qualifiers

1. .11932

/organism="Neisseria meningitidis MC58"

/strain="MC58"

/db\_xref="taxon:122586"

/note="serogroup: B"

102. .1097

/gene="NMB1573"

102. .1097

/note="similar to GB:M34930 SP:P21302 GB:X64860 PID:150241  
 PID:44847 percent identity: 97.58, identified by sequence  
 similarity; putative"

/transl\_table=11  
 /codon\_start=1

/product="ornithine carbamoyltransferase, catabolic"

/protein\_id="AAF41926.1"

/db\_xref="GI:7226821"

/translation="MNKNRHFELKLPFPEETAYVLDLAEKAKKACREIORMKG  
 KNILFEKISTRTCAFEVAAADQAGVYLEPSAQIGHKESIKDTAVLGRMDA  
 IERFGQEVVEELAKYAGVFNELNREHPDMLADALMREHSGKPLNQTAVAY  
 GDARYNMGNSLLILGAKLMDVRIGAPOSIMPSEGIIAAHAARKETGAKITLENH  
 EAVKNDPEIHTDVMVSGEPREYQERIDILKQRYVPELMAASGNPOVFMHCIPAF  
 HNRETKYGEWTEFEGLVGEVPEEIRSPASIVDFDAENRMHTIKAVYVALGD"

complement(1321. .2334)

/gene="NMB1574"

complement(1321. .2334)

/note="similar to GB:AE000512 percent identity: 77.25;  
 identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ketol-acid reductoisomerase"

/protein\_id="AAF41927.1"

/db\_xref="GI:7226822"

/translation="MVOYDKDADLSIKGTVAITIGVSGGHANANLKDQSVNVI  
 GARGSWKKAAGHYKVTVAEKTRADYVMLLPETRPAYHAETVANLEGATL  
 AFAGFVHTNQIVPRADLVINAPRGPCHIVKSEKRGGSPLIAYVQDSSGAK  
 DIALSTVAANGTGGVLETTFRETEDELRQAVLCGGVELIKAGETITLGAAG  
 PEMAYFCELEMKLIDLIFEGEIANNVISINNAEGEYVTGPEVNVSSKRAMRA  
 LKRIOTGEYAKMFIQEGNVNVYASMTARRRLNADHGVKGAQQLRAMMPVTANKLVQD  
 KRN"

complement(2414. .2707)

/gene="NMB1575"

complement(2414. .2707)

/note="conserved hypothetical protein; identified by  
 Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAF41928.1"

/db\_xref="GI:7226823"

/translation="MSNIKIVALTAKPEVTEITLAAQFKELVAKSRAEGNTSYDHLQ  
 EICKPNFVFEVWKSQAATDEENASHHQAFVQSDGKTEALEITLMEVAD"

complement(2774. .3265)

/gene="NMB1576"

complement(2774. .3265)

/note="similar to PID:147816 GB:U00096 PID:1786266 percent  
 identity: 79.75; identified by sequence similarity;  
 putative"

/codon\_start=1

/transl\_table=11

/product="acetylacetyl synthase III, small subunit"

/protein\_id="AAF41929.1"

/db\_xref="GI:7226824"

/translation="MRHILSVLIENSGAMSRVYGLFSARDYNIDSLAVAPTEKTLIS

gene

CDS

complement(3276. .5003)

/gene="NMB1577"

complement(3276. .5003)

/gene="NMB1577"

/note="similar to GB:142023 SP:P45261 PID:1007821  
 PID:1221728 PID:1205817 percent identity: 74.51;  
 identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="acetylacetyl synthase III, large subunit"

/protein\_id="AAF41930.1"

/db\_xref="GI:7226825"

/translation="MQLSAQIITVSLAKAEYVEYFGGAVIEIYDALFQNLNFKH  
 ILTRHQAQAAVHADAVARVSGKVALVTSQGVNALGIAATYDTSIPNVISGVQV  
 GNSLIGTDAFOEVDIVGTRPCVKNHELVTDINELETIKAKAQAASGRPGPVVVD  
 PKDVTQAMAKESYPOEDIFIRSYOPVOGHITGOIKKAVOMLAKRPVYFGGVVIG  
 NASEELTRFVRMTGAPCTGTLMGIGAVPSGDQPGMIGMGTGTENANLQMONADVLA  
 VGARPDQVSVSPKPEKAKYTHIDVPSSIAKRVKIDIVGVKAVLSEKVALM  
 OKQESVPSDALCKMKKTIEBMSKDCMPFDGSETIKPOVYIQKLAETVTSALITS  
 DVQOHQFAQYYPFERPROMNSGLGTMGVLPAIGAKLAADPODFCTTGSGSI  
 OMNIQELSTCFQYRIPVNYITLNNGLGWVOMQETIYGGRESEYFDSLPPFVKAE  
 AVGHIGIRVDKSDVGCALLEALNQKDRLVFIIDFLTDQKQVMPVWGNGKIDENVLP  
 PHMRADGKA"

complement(5763. .5847)

/gene="trna-leu-4"

complement(5763. .5847)

/gene="trna-leu-4"

/product="trna-leu-4"

6124. .6777

/gene="NMB1578"

6124. .6777

/gene="NMB1578"

/note="conserved hypothetical protein; identified by  
 Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAF41931.1"

/db\_xref="GI:7226826"

/translation="MFSVRSFLPLPGFVLAALAAACKPODNSAAQVASSASAAENA  
 AKRQGTMRKEDIGDETLTGEGGKPNISDLGKVVYIISFGTHCPDVCPELIT  
 YSDTLKQGGQKDKVYVPSYIDPRDPEITIGKAKQFNPFISLTATGCONLPYIK  
 OQRYVSAVANKDSENLVDHSSGAVLIDKNGEVALFSPYSEPEITADVRLL"  
 6856. .7509

/gene="NMB1579"

6856. .7509

/gene="NMB1579"

/note="similar to GB:AL009126 percent identity: 67.89;  
 identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ATP phosphoribosyltransferase"

/protein\_id="AAF41932.1"

/db\_xref="GI:7226827"

/translation="MODNALTALSKGRFERTLPDLAAGIVPTEPEKSRKLLIGT  
 NHEINRLVIRATIDVPTVYRGADFGIAGKVLLEHGGTGLYRPLDLEIAKCRMVA  
 VKRGDFYEAASQPGRLKATKVPYIAASHFGKGVHVDIKLYGSMELAGVYISDA  
 IVDLVSTGNTLKAAGLEAVEHIVDISRLVYVKKALKTKYALLEPIIQAQGAAYAK"  
 7613. .8512

/gene="NMB1580"

7613. .8512

/gene="NMB1580"

/note="hypothetical protein; identified by Glimmer2;  
 putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAF41933.1"

/db\_xref="GI:7226828"

/translation="WVAKIKRKSDSTLSVLNNGRRRFPVYVCTGDIKKDKITLYIGKCGG



NR1FEHFWASRSODPVSGEITIDKIKAKICKKIGRIITISYLHLEVEALAAASALIH  
 FKSVLGKKKKKTAGHGPISIEEDRREGSESLPNEINPGLAIKINAPDL  
 DTDELLDYLEDNODDANIKSRITLGNWYKDVASKVITGVHTGTONAAVSAVEDG  
 FEIMVEETKNGKRSRYRFRFTSRSEEVKALGLQOKCELPKFGSGGEKAVIRPKTE  
 TETEOENIOTTPNPKIKKEKTKS"  
 8509. 9798  
 /gene="NM01581"  
 8509. 9798  
 /gene="NM01581"  
 /note="similar to GB:A000657 percent identity: 71.03;  
 identified by sequence similarity: putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="histidinol dehydrogenase"  
 /protein\_id="AA041934.1"  
 /db\_xref="GI:726829"

Query Match 8.5%; Score 90.8; DB 1; Length 11932;  
 Best Local Similarity 47.7%; Pred. No. 2.2e-14;  
 Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

QY 182 AAGCGTTCTCGGATTTTTCGGGCTAAGAGAAATTTAGTCTCGGTACGGTT 241  
 DB 10058 AAGCATTTAGTGGGCTTGCAGATTTCCCGATGCGCCGACATCCGCTGGGCAACGGTT 10117  
 QY 242 CGAGCAACTATATCTACTCTCATATAGTATAGTGAAGTATTAACATACCGCTTTACA 301  
 DB 10118 CGAGCAACTATATCTACTCTCATATAGTATAGTGAAGTATTAACATACCGCTTTACA 301  
 QY 302 TACCTTTCCGACCTTCCCATGTAGAGATATAGTCCGAATTTCTCGGAAGACCCCTCG 361  
 DB 10178 TACCTTTCCGACCTTCCCATGTAGAGATATAGTCCGAATTTCTCGGAAGACCCCTCG 361  
 QY 362 TAAAGTTCACTGAGAGAACTTGTATATAGTATTAAGAAATTTAGTATTAATAG 421  
 DB 10238 TCGGCGTTCCATCACTGAGAGATTTACCTTACCTGCGCCGCTGCTGCAAGCCGTCA 10297  
 QY 422 AGAAGAAAAACCGCTTCTCGGCTTGTATACCAACCAACCCGAGGAAACCTCT 481  
 DB 10298 GGAAGAACCGCCCTGCGCTTATGCGCTACCCCAACACCCGAGGCTATGCT 10357  
 QY 482 TTTTCCAGGGAAGATTGA---GGAATTAAGAACAGGGGTTTCTGTGTAATAGAG 538  
 DB 10358 TCAAGGCTGCGGAAATCGAAGCGCTCATGCAAGCTTCAAGCGGATCTGTGCTGATG 10417  
 QY 539 AAGCTTACTATCATTTCTCGGAGAAACCTTCTGGAAGACCGGCTCAAAAGGAGATA 598  
 DB 10418 AAGCTTACTATCATTTCTCGGAGAAACCTTCTGGAAGACCGGCTCAAAAGGAGATA 598  
 QY 599 CGGTAGTTTGTAGGACACTTTCAAAATCGTATGCGGATTTAAGGTTAGGGATTTTAA 658  
 DB 10478 TGTATAGCTTACGACCTTCAAGAAATCGGTTTTCGCGACTGCTATCGGTTATGCGG 10537  
 QY 659 TAGGAGAGGGGAAATCGTCTCAAAATTTAAGAGTGAAGTACCTTCAAGTACCT 718  
 DB 10538 CAGGCTGCGCCGGAAGTATCGGCAAAATTCCTGCGCCCTTCAAAATATACAC 10597  
 QY 719 ACCCTCTCGAGTATGAGCAAAAGTTCTCTCAAGAGAGGAGAAATTTCTTAATGAAA 778  
 DB 10598 AATTGAGCTTACGACCTGCAAACTCGCTGCGGCACTAGGCAATTAATCTC---GCCA 10654  
 QY 779 AGATACAGAGAGTTGTATACAGAGCGAAGATGTATGAGCAATTAAGCAAAATTAAGAG 838  
 DB 10655 ACATTCGACAGCTTGAAGAAAGCAAGCGAGATGTTCGCGCAATTTGGCAAAATATGCG 10714  
 QY 839 GAGTTGAGGTTTTTCCAGATAGGCTACTTCTT 872  
 DB 10715 GTCTGAACACCTTTTCAAGTTCAGGCAAACTTCAAT 10748

RESULT 7  
 AX044033 349980 bp DNA linear PAT 24-NOV-2000  
 LOCUS  
 DEFINITION  
 AX044033 Sequence 112 from Patent WO006791.  
 ACCESSION  
 VERSION  
 AX044033.1 GI:11342917  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria meningitidis.  
 Neisseria meningitidis  
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
 Neisseria.  
 1 (bases 1 to 349980)  
 Pizzaz, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,  
 Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarcell, M.,  
 Scarlato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.  
 TITLE  
 Neisseria genomic sequences and methods of their use  
 JOURNAL  
 Patent: WO 006791-A 112 09-NOV-2000;  
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
 FEATURES  
 source  
 1. 349980  
 /organism="Neisseria meningitidis"  
 /db\_xref="taxon:487"  
 /note="sequence too long, cut in 8 pieces, -seq 1: 1 to  
 349980 349980 bases -seq 108: 300001 to 649980 349980  
 bases -seq 109: 600001 to 949980 349980 bases -seq 110:  
 900001 to 1249980 349980 bases -seq 111: 1200001 to 1549980  
 349980 bases -seq 112: 1500001 to 1849980 349980 bases -seq  
 113: 1800001 to 2149980 349980 bases -seq 114: 2100001 to  
 2272325 172325 bases"

BASE COUNT 87189 a 93501 c 84627 g 84663 t  
 ORIGIN

Query Match 8.5%; Score 90.8; DB 6; Length 349980;  
 Best Local Similarity 47.7%; Pred. No. 3.1e-14;  
 Matches 331; Conservative 0; Mismatches 357; Indels 6; Gaps 2;

QY 182 AAGCGTTCTCGGATTTTTCGGGCTAAGAGAAATTTAGTCTCGGTACGGTT 241  
 DB 144619 AAGCATTTAGTGGGCTTGCAGATTTCCCGATGCGCCGACATCCGCTGGGCAACGGTT 144678  
 QY 242 CGGAGCAACTATATCTACTCTCATATAGTATAGTGAAGTATTAACATACCGCTTTACA 301  
 DB 14479 CGGAGCAACTATATCTACTCTCATATAGTATAGTGAAGTATTAACATACCGCTTTACA 301  
 QY 302 TACCTTTCCGACCTTCCCATGTAGAGATATAGTCCGAATTTCTCGGAAGACCCCTCG 361  
 DB 144739 TACCTTTCCGACCTTCCCATGTAGAGATATAGTCCGAATTTCTCGGAAGACCCCTCG 361  
 QY 362 TAAAGTTCACTGAGAGAACTTGTATATAGTATTAAGAAATTTAGTATTAATAG 421  
 DB 144799 TCGGCGTTCCATCACTGAGAGATTTACCTTACCTGCGCCGCTGCTGCAAGCCGTCA 144858  
 QY 422 AGAAGAAAAACCGCTTCTCGGCTTGTATACCAACCAACCCGAGGAAACCTCT 481  
 DB 144859 GGAAGAACCGCCCTGCGCTTATGCGCTACCCCAACACCCGAGGCTATGCT 144918  
 QY 482 TTTTCCAGGGAAGATTGA---GGAATTAAGAACAGGGGTTTCTGTGTAATAGAG 538  
 DB 144919 TCAAGGCTGCGGAAATCGAAGCGCTCATGCAAGCTTCAAGCGGATCTGTGCTGATG 144978  
 QY 539 AAGCTTACTATCATTTCTCGGAGAAACCTTCTGGAAGACCGGCTCAAAAGGAGATA 598  
 DB 144979 AAGCTTACTATCATTTCTCGGAGAAACCTTCTGGAAGACCGGCTCAAAAGGAGATA 598  
 QY 599 CGGTAGTTTGTAGGACACTTTCAAAATCGTATGCGGATTTAAGGTTAGGGATTTTAA 658  
 DB 145039 TGTATAGCTTACGACCTTCAAGAAATCGGTTTTCGCGACTGCTATCGGTTATGCGG 145098  
 QY 659 TAGGAGAGGGGAAATCGTCTCAAAATTTAAGAGTGAAGTACCTTCAAGTACCT 718  
 DB 145099 CAGGCTGCGCCGGAAGTATCGGCAAAATTCCTGCGCCCTTCAAAATATGCAAC 145158



QY 719 ACCGCTCTGAGTATGGCAAAAGTTCTCTCAGGAGGAGAGAGATTCTTAATGAAA 778  
 Db 145159 AATTGAGCCTGACCACTGCCAAACCTGCGGACATGACATTATCTCT--GCCA 145215  
 QY 779 AGATACAGGAGGTTGTAACAGACGAGAAAGATGTACGACGAATGAAGAAATGAAG 838  
 Db 145216 ACATTCACAGACCTGAAACGAGACGAGATGTTCGCGCAATTGGCAAAATATGCG 145275  
 QY 839 GAGTGTAGGTTTTCGAGTAAGGCTAAGCTCTT 872  
 Db 145276 GTCTGACACCTTTTCAAGTCAGCAAACTTCAT 145309  
 RESULT 8  
 AE001764/C 20334 bp DNA linear BCT 02-JUN-1999  
 LOCUS Thermotoga maritima section 76 of 136 of the complete genome.  
 DEFINITION AE001764 AE000512  
 VERSION AE001764.1 GI:4981561  
 KEYWORDS  
 SOURCE Thermotoga maritima.  
 ORGANISM Thermotoga maritima.  
 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 REFERENCE 1 (bases 1 to 20334)  
 AUTHORS Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., Fraser, C.M. et al.  
 TITLE Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima  
 JOURNAL Nature 399 (6734), 323-329 (1999)  
 MEDLINE 99287316  
 PUBMED 10360571  
 REFERENCE 2 (bases 1 to 20334)  
 AUTHORS Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source location/Qualifiers  
 1..20334  
 /organism="Thermotoga maritima"  
 /db\_xref="taxon:2336"  
 complement(70..1338)  
 /gene="TM1021"  
 complement(70..1338)  
 /gene="TM1021"  
 /note="similar to GP:2695718 percent identity: 100.00;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="permease, putative"  
 /protein\_id="AAD36098.1"  
 /db\_xref="GI:4981562"  
 /translation="MAVFTILMLVLMADQWVMSNIGALIEQENITDQIGLVASS  
 FTVIGALVSLWGYLADRYSRKMLIYSLVGEIPCLMSAFSRSGELFPMRALTGIG  
 VGSAPFIVYSMDGMDVGRKGVALLISSAIGSLVGLMIVGFLGPKYGRVPEIA  
 VSVNIFFAVLISIFLKEPKRGAFEGIGELVOSGYEYKAPLSDYAKLVKTNLL  
 LPEQGIAGTIPMGCAIPYPLVEFRREREGISVETATVVELVPGGNVAGITIGLNGAS  
 IYAKSRPLPLGCSITATGCTFTVTITLDYMSLVLMILGLTASTFASLTGPNVKEM  
 LTNVNEPQERGRFISFTNLTDSIGTEFGKFAAGVMSVALSGAALKVSAFWMLICAV  
 LLEFVETFAKDEVERLQKTMIELAKNSQTR"  
 complement(1341..2102)  
 /gene="TM1022"  
 complement(1341..2102)  
 CDS

/gene="TM1022"  
 /note="similar to GP:2695717 percent identity: 100.00;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="esterase"  
 /protein\_id="AAD36099.1"  
 /db\_xref="GI:4981563"  
 /translation="MNFPRCKTYVKKSLPIFLEGNGEVLFIHGYSQPHDEYMAKEY  
 NRAFTVSVPRLPGHCGDDELTTTARDMLRAFDAYYDLKAIQDRYVYVGI.SMGV  
 IALLASOMNPKRYVTIAATFVFDKIVITLPIKLFTKMPCEMTEKEDPIETLR  
 KEYSTYMPKQDAELIKLMLKLAARSVKISATLVAAKNDNNVPMKAEFIYNNTRS  
 EKKRLVFEKSGHVLSDVEREDVTAVIEMLKEG"  
 2200..2721  
 /gene="TM1023"  
 2200..2721  
 /gene="TM1023"  
 /note="similar to PID:992558 percent identity: 53.12;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAD36100.1"  
 /db\_xref="GI:4981564"  
 /translation="MKRSPLVTLILMLIGLVFESSOPPDVSGROSNVYKFLKVD  
 NVLDFQTRWYRNILRSILEKWPDPKPGCEDLVRSAGFLYFINGILISFPESTYTL  
 RKYVFSILMGVSLPTLLAVDEVNOSFRGSGASLYVIYDMKAVIGTILFLILLTL  
 LKRRRLRRRLIL"  
 2790..3251  
 /gene="TM1024"  
 2790..3251  
 /gene="TM1024"  
 /note="similar to percent identity: 0.00; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAD36101.1"  
 /db\_xref="GI:4981565"  
 /translation="MEKEVFLLEVSTFVYGDERSPRIRKFPIDIEVEYVYPTS  
 KKVDYVMISSNKPSPKIEGDTIILSDDEPQEGDLVAVLKDRTIVGRVKEQED  
 IYVEYGEVGIYRCPFIYDQLRKSSELSFNPYLAKITAILRKRPAA"  
 3309..3407  
 /gene="TM1025"  
 3309..3407  
 /gene="TM1025"  
 /note="similar to percent identity: 0.00; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAD36102.1"  
 /db\_xref="GI:4981566"  
 /translation="MKQKPSPLWPERKLTNENLARKNHVDFEL"  
 4040..5248  
 /gene="TM1026"  
 4040..5248  
 /gene="TM1026"  
 /note="similar to GB:I77117 SP:Q58161 PID:1591465 percent  
 identity: 55.70; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transposase, putative"  
 /protein\_id="AAD36103.1"  
 /db\_xref="GI:4981567"  
 /translation="MPRVIVRYTKLAIPGHNLQICEELNRIARIVYKTSYNOKIHO  
 KKGWLSWPTADKYILRAENIKIHVSKQAFQOLYQALQGYKRAKKKQDAPPK  
 KRYVLPFWKSAKILSDGTLRSLGKREKPEYVQIPLRPILIKARIVLFGEGYLL  
 HLATEVELEKSSANSQVAAVDGLVLRPLTCDGKEVVISYGGVLSVLRNNRKLASL  
 OSLAECKGSRRYTNKLIVRAKRVLRKLNQINDIMKRTSSFGICLRQIGTIVG  
 DVATIRERADYSNANOKIHQWQPRKLIEMIRYKAQDFGIEVKLISPAKSKCPVGC  
 AKNKPNGRRYHCKTGGFEYHRDVGGAIVIMKRYGTGGVYAGLAIPVAGVAFPHILCGH

gene  
GASLAPWKVA  
5309..5644  
/gene="TM1027"  
5309..5644  
/gene="TM1027"  
/note="similar to percent identity: 0.00; identified by sequence similarity; putative"  
/transl\_start=1  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="A036104.1"  
/db\_xref="GI:4981568"  
/translation="MSSEFLGOGNARYSATYSMGTVPKAKELIGFSPISGFIHPDLKDVLEIPDIFRWQNEPQSYGVSVIYIFNGECRACSCYENGAAAEIKTYOTGLVTVQOLVE"  
5922..6803  
/gene="TM1028"  
5922..6803  
/gene="TM1028"  
/note="similar to GB:AE000782 percent identity: 60.48; identified by sequence similarity; putative"  
/transl\_start=1  
/codon\_start=1  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="A036105.1"  
/db\_xref="GI:4981569"  
/translation="MSVVEYRDLTKYVGRGVGVTFSEGESEIFGPIGNGAKTTIRLLGLIPDSGLAEITFGKDYKKEKTRKRVGIIPEGVSFPEVTVPEELRYASFYKRVMDVYKELGVESLDTKRIKELSGNKKKVAIVQALHRRKLLIDEPITGLDPIVQNTVFELKNEKEKGTTFVSSHIISSEVERLIDRVAMIKDGR1IREKVENLNGERRVAVKGENLEKLDLAENRKLKFNENGTAFLEPSGNERLLIELRLKLSDFWEERSEEFMSYREGE"  
6800..7591  
/gene="TM1029"  
6800..7591  
/gene="TM1029"  
/note="similar to SP:P42333 percent identity: 50.27; identified by sequence similarity; putative"  
/transl\_start=1  
/codon\_start=1  
/product="ABC transporter, permease protein, putative"  
/protein\_id="A036106.1"  
/db\_xref="GI:4981570"  
/translation="MNIFFMDMKRYKSTIAWTVLLIOLMTAAFPYPSNAKTEETITKMKVMPKAFKLEGLDMDSNIMNYLAWSIYVTVGVGASIVGVSVEKEENEKTVRELISKPEVREVIYSKFEESLVHLLFDALFSLMFANITSSSPVSEKEVYFASQVFLHTLNMISFVGTIRSDGALSTIGIVFVLYLNMISKITDKAEFLKFTPEFYADPSVIRHGFEPFSGLEFALVINALIISIVLFSKDLIV"  
7602..8204  
/gene="TM1030"  
7602..8204  
/gene="TM1030"  
/note="similar to GB:S87512 SP:P43506 PID:623036 PID:247052 percent identity: 48.00; identified by sequence similarity; putative"  
/transl\_start=1  
/codon\_start=1  
/product="transcriptional regulator, TetR family"  
/protein\_id="A036107.1"  
/db\_xref="GI:4981571"  
/translation="WLSKRDAILKAAYEVEKKGDRATDEIAEKAVAKGLIFHYFKNEELVQAYMSVTEKLOKEFENELMKNRNDIDFEMERTEKLEYSASHPEEAFLLITVSVDEGLRKRLIDLEKSORVEFDLEKLDLADVETETALAKFLWFFSGFEEVYLRTYOGKPELKRDMNTLVEEYVLMRLIKKGMK"  
complement(8229..8471)  
/gene="TM1031"  
complement(8229..8471)  
/gene="TM1031"  
/note="similar to GB:M60116 SP:P23171 PID:144907 percent identity: 73.61; identified by sequence similarity;  
Query Match 8.4%; Score 89.2; DB 1; Length 20334;  
Best Local Similarity 54.3%; Pred. No. 6.7e-14;

Matches 225; Conservative 0; Mismatches 183; Indels 6; Gaps 2;  
QY 427 GAAAAACCGTTCGGGTACTTTCCTTACCCCAACACCCAGGGAACCTCTTCC 486  
DB 15718 GGAGAGGAGACGCTGTTTCACTCCGAACCCGACATCCACAGGGCCATCTTGAA 15659  
QY 487 AGGGGAAAGATTGAGAGATAAGAAACAGGGGCTTTCTGTATATAGCAGAACCTAC 546  
DB 15658 AACAGGAAATGAAAGATCTCGAAACGGGTGCTTCGTGCGCTGAGAACCTAC 15599  
QY 547 TATCATTTACTCCGGAAGAAACCTTCTGGAAGACGGCTCAAAAGGAATACGTTAGT 606  
DB 15598 TACGAATTCACAGGAAAGATATGTATTTCTGAAGAAATACGAATCTGCTGTG 15539  
QY 607 TTGAGACACTTTCAGAAA--ATCGTATGCGGACTTTAAGGATTAAGGATTTAATAGG 663  
DB 15538 ATCAGACACTTTCGAAACCGTTCCTCGGAGGCAACGTTGATGTTGTGCGC 15479  
QY 664 AAGGGGAAATGCTCTCAGAAATTACAGATGACACTCCTTGAAGTACGACCTACCC 723  
DB 15478 TCGGAGAACCTTATGACCTTACACAGGGTACGACTTCTTAACGTGAGCTACGTC 15419  
QY 724 TCTCAGCTGATGGCAAAAGTTCTCCACAGGAGGAGAAAGAAATCTAATGAAAGATA 783  
DB 15418 TCCAGATGTTCCAGAGTACCTCTC---GATCACAGAGATCTTGAAGAAAGAACG 15362  
QY 784 CAGAGGTTGTACAGACGAGAAAGATGTACGACGAAATGACGAATAAGTAAGAA 837  
DB 15361 AATTTCATCTCGAAGACGAGATGAAGATGCTCTCAGGAAATGGA 15308  
RESULT 9  
NMA622491 326301 bp DNA linear BCT 04-DEC-2000  
LOCUS  
DEFINITION  
Neisseria meningitidis serogroup A strain 22491 complete genome;  
segment 6/7.  
ACCESSION  
AL162757 AL157959  
VERSION  
AL162757.2 GI:7380371  
KEYWORDS  
Neisseria meningitidis 22491.  
Neisseria meningitidis 22491.  
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
Neisseria.  
REFERENCE  
1 (bases 1 to 326301)  
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltham,D., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Raftery,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.  
Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491  
JOURNAL  
Nature 404 (6777), 502-506 (2000)  
MEDLINE  
20222556  
PUBMED  
10761919  
REFERENCE  
2 (bases 1 to 326301)  
Parkhill,J.  
Direct Submission  
Submitted on behalf of the Neisseria  
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
COMMENT  
Notes:  
Details of N. meningitidis sequencing at the Sanger Centre are  
available on the World Wide Web  
(URL: [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).  
FEATURES  
source  
1..326301  
/organism="Neisseria meningitidis 22491"  
/strain="22491"  
/db\_xref="taxon:122587"  
/note="serogroup: A"  
196..311  
/note="REP 2; hmfs hit to HMM REP 2 (1 - 109), score:

```

repeat_unit      87.43"
                  434. .453
                  /note=">= 90% match to ATCCCNNNNNNGGAAAT"
repeat_unit      /label=dRS3
                  498. .560
                  /note="Correia element; hmms hit to HM Correira (1 - 62),
                  score: 78.71"
                  /label=Correira
                  562. .604
                  /note="Correia element; hmms hit to HM Correira (114 -
                  156), score: 51.92"
misc_feature      /label=Correira
                  complement(667. .676)
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(668. .1813)
                  /gene="dape"
                  complement(668. .1813)
                  /EC_number="3.5.1.18"
                  /note="NMA1730, dape, succinyl-diaminopimelate
                  desuccinylase, len: 381aa; similar to many eg. SW:P24176
                  (DAPE_ECOLI) dape, succinyl-diaminopimelate desuccinylase
                  from Escherichia coli (375 aa) (fasta scores; E(): 0, 56.98
                  identity in 371 aa overlap. Contains Pfam match to entry
                  PF01546 Peptidase_M20."
                  /codon_start=1
                  /transl_table=11
                  /product="succinyl-diaminopimelate desuccinylase"
                  /protein_id="CAB84958.1"
                  /db_xref="GI:7380372"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MTETQSLELAKALISRPVTPDDRCOKLIVERLYIGFAAEEL
                  HFQTKNIMWRKGTQVPCFAGHTDVPTGPKVEMDSPEEPERDGRYGGGAAM
                  KTJACFVTCERFAVAHPHOGSIALILTSDEGDLDTTVVLYKARGELIDIC
                  ICEPTAVDKLGMKIKNGRSLSGNLTAKGSHIAYPHLAINPVPKAPALLEIIO
                  EWDEGKNKYPPPTSFQISNTNGCGANVPIGELNKNFNRSTESTEACLRVHAI
                  LDKRGVYDIOMSSGQPTLTQAKGLTDVAKAIATETGIEALESTTGGSDFRTKA
                  IAKELIPLSPSNTIHOINENVRDIPKLSAVYEGILARLAGNAV"
misc_feature      complement(836. .1786)
                  /gene="dape"
                  /note="Pfam match to entry PF01546 Peptidase_M20, score
                  252.50, E-value 5.7e-72"
                  complement(1863. .1872)
misc_feature      /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(1927. .2547)
                  /gene="NMA1731"
                  complement(1927. .2547)
                  /gene="NMA1731"
                  /note="NMA1731, conserved hypothetical protein, len:
                  206aa; similar to SW:P44013 (Y552_HAEIN) hypothetical
                  protein from Haemophilus influenzae (207 aa) (fasta scores;
                  E(): 0, 59.68 identity in 203 aa overlap."
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="CAB84959.1"
                  /db_xref="GI:7380373"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MLTPKSCDLFNIPFOFSQTKKYPESIPQIKADYKENNOIWO
                  LIOOVAELGAPAPAPHIEMKMGQVRAHFAFPEYKQYKNAALITLILNRRISV
                  SLDMHCYKADVPALPDYNRMLDNPTETKYSFDMHGHESEYDYRTVVAQONESDR
                  KLONDEDFCTIGHIERDDLGRODVAKWIAETVEDLLPLYEACHGK"
RBS              complement(2551. .2558)
misc_feature      complement(2570. .2579)
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(2581. .3090)
                  /gene="NMA1732"
                  complement(2581. .3090)
                  /gene="NMA1732"
                  /note="NMA1732, conserved hypothetical protein, len:

```

```

169aa; similar to SW:P43971 (Y243_HAEIN) hypothetical
protein from Haemophilus influenzae (172 aa) (fasta scores;
E(): 0, 49.7% identity in 169 aa overlap."
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="CAB84960.1"
                  /db_xref="GI:7380374"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="MNPFTQSVTFAPEIEMLYACHKVRFCGGVAMLSDYAENG
                  CNOVLQITQITRYFVNAAPLHHEDEENPEPLLDYAPAOAESVDLLRQVYSHGN
                  MDAAVEFAKLEADNAVIPDAEFKRFVAGYDVHIALIEEPLDMKMFIPKEXLTG
                  EIMARRRR"
                  complement(3185. .3736)
                  /gene="NMA1733"
                  complement(3185. .3736)
                  /gene="NMA1733"
                  /note="NMA1733, H.8 outer membrane protein, len: 183aa;
                  strongly similar to SW:P07212 (H8_NHME) H.8 outer
                  membrane protein from Neisseria meningitidis (183 aa)
                  (fasta scores; E(): 0, 96.7% identity in 183 aa overlap.
                  Contains Pfam match to entry PF00127 copper-bind, Copper
                  binding proteins, plastocyanin/azurin family and Prosite
                  match to PS00196 Type-1 copper (blue) proteins signature."
                  /codon_start=1
                  /transl_table=11
                  /product="H.8 outer membrane protein"
                  /protein_id="CAB84961.1"
                  /db_xref="GI:7380375"
                  /db_xref="SWISS-PROT:P57025"
                  /translation="MKAYLALISAAYIGLAACSOEPAPAAEPAPAEAPASAPAE
                  AAPADAAPAGNCATVAYESNDNQENFKDIQVACKKEFTITLKHGTQPKSMGH
                  EIVIGTCTMDGIFPKDVGAAAPDYKPDAAVVAHTKLIGGEESLTLDPKALADG
                  EKFAKTPFGHGLAMKXTLVD"
                  complement(3191. .3568)
                  /gene="NMA1733"
                  /note="Pfam match to entry PF00127 copper-bind, Copper
                  binding proteins, plastocyanin/azurin family, score
                  218.60, E-value 9.1e-62"
                  complement(3212. .3253)
                  /gene="NMA1733"
                  /note="PS00196 Type-1 copper (blue) proteins signature."
                  complement(4017. .4439)
                  /gene="NMA1734"
                  complement(4017. .4439)
                  /gene="NMA1734"
                  /note="NMA1734, unknown, len: 140aa"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB84962.1"
                  /db_xref="GI:7380376"
                  /db_xref="SPTRMBL:09JUL0"
                  /translation="WNLQDLGRDALLSGLLKQACQMRDLDAVAKLLPANLHPHQR
                  ACIEDGRVLTLAANNMAASRLKIMAPSVLPQAGIDASIRSVSVLVPKPKPKPTNT
                  IHLSKAALSFCSAAVVKLEKRHPDELALANLVRHGA"
                  4306. .4315
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(4486. .4495)
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  /label=DUS
                  complement(4570. .7320)
                  /gene="seca"
                  4570. .7320
                  /gene="seca"
                  /note="seca"
                  /note="NMA1735, seca, preprotein translocase SecA subunit,
                  len: 916aa; similar to many eg. SW:P10408 (SECA_ECOLI)
                  preprotein translocase SecA subunit from Escherichia coli
                  (901 aa) (fasta scores; E(): 0, 58.0% identity in 920 aa
                  overlap. Contains Pfam match to entry PF01043
                  SecA_protein, SecA protein, amino terminal region and
                  Prosite match to PS01312 Protein seca signatures."

```



[illegible]

					/transl_table=11 /product="alcohol dehydrogenase, iron-containing" /protein_id="AAB89231.1" /db_xref="GI:2648512" /translation="MRVYVGGRLEIRIEEYERLKGKGVVFIAARAVAKAEINLSF DYIESDVEPEPEIGTIVDDIIEKSFEDTVYVGGSSTIDPAKLTVAAAGEKAVEIL GKKIPERKLLICTTTTGTSSEYTKLAVERIKPRKVNYVPDSDSLADAI IDPKLTLL LSCPHTVASSGIDAFCHALEAYSLFSNPISDMFAEKVAIIIPKALREAYANENR AKTDWISFASLLAGIAFNNSAGTGLGHALGYAHNHINSPHGSSVGTITMPYLYQYNAID LEKHRIARALGESSSLRETALRAGIICFPALKDKLGMPHKLSDVGAGEDDIETIVDR IFLSKHVRNRPVRYREEMDVYRKSIHGILLREBY"
gene					complement(6502..7077) /gene="Af2020" complement(6502..7077) /gene="Af2020" /note="hypothetical protein; identified by Genemark; putative"
CDS					/codon_start=1 /transl_table=11 /product="A. fulgidus predicted coding region Af2020" /protein_id="AAB89241.1" /db_xref="GI:2648523" /translation="MVKLRHTTVSSETARKIEDLKRRHTSSSVKEAVDLVTSEN FSRLGDEDLIILAFIRELNELCAKDHTALVEDGADEARVESMIEMAIVYLKSPIS LDDEEELSVMVARVLNNLRHAHEVOCKGKLNLFVYIHDMRSKAIVSELHNILKITYL EKSKYKYEAMOVDTITNGFSVLPFKPDSD"
gene					7057..8148 /gene="Af2021" 7057..8148 /note="similar to GB:DJ77799.SP:P39763.PID:520840 PID:1377825.GB:AL009126.percent identity: 26.60; identified by sequence similarity; putative"
CDS					/codon_start=1 /transl_table=11 /product="rod shape-determining protein (mreB)" /protein_id="AAB89230.1" /db_xref="GI:2648511" /translation="MEPDPYHGFPPTPLKTYPEIFYIGKCTENMPQVGLDGTNT KITADGNKLVLPSTIVAYGEKMSLGETKRSYIGDALSTIVOSMENYVAILPIHG RHITSYIELAKYGLINKLKYDGLIANGCYSKKRRELKRELBESLNSKIILPE PVGMAYMGMETGCYVDIGRTDVIYAIDMETIKDTHMLGDVKTLENLEVNRNI GISITPEBMKLLTFEBEGLETGRIGRGKVSAREILPEZEGYKWKVERIASRVKLY LEGISTFLVENFALTGGSMIPGVYENMKFSFDIGDKRPDEPIKANSTIGYILAKS LVVAETEVSQSDAAILDADDKKKRGKKGK"
gene					8145..8672 /gene="Af2022" 8145..8672 /gene="Af2022" /note="hypothetical protein; identified by Genemark; putative"
CDS					
Query Match	8.1%;	Score 86;	DB 1;	Length 22014;	
Best Local Similarity	47.4%;	Pred. No. 5.5e-13;			
Matches	476;	Conservative	0;	Mismatches 495;	Indels 33; Gaps 6;
OY	61	GTCAGCGTTCCCTTAACGAATTCGCCCTACGACTTCCCAGAGAGATAAACAAGGACC	120		
Dd	9600	GTCACACTTCTTCCAAGAGAATCCTTATGACCGCACGACGAGGTATTAGGAGGCTTAC	9659		
OY	121	TTAGAAGAAATAAAAAGCTGCCCTTGACAAATACCCAGACCCCGAAGCAAAGAGTTA	180		
Dd	9660	CTTAGAGCTGTGAANAATCATTTCC-----AAGATATCCTTAAGCGGACTACGCCGCTG	9713		
OY	181	AAAGCGGTTCTTCGGGATTTTTTCGGCGTTAAGGAAGAAAATTATGTTCTCGTACGGT	240		
Dd	9714	AAAAAGGCCCTTGGACAGATACACGGGCTTTGAATTTGAAAAACATACGCTGGATGCGA	9773		
OY	241	TGCGACCAACTATATACCTCTCAATATAGCTTATAGGTGAACCTTATACATACCGGTTAC	300		
Dd	9774	GCATCTGAGCTAAATTTCTGCTGATCGAACGTTCTGCTGGAAAGAGCTGACAGGTTGTC	9833		
OY	301	ATACCTGTCCACCTTCCCATGTAGAGATAGTGCMAAAGCTTCTCGAAGACCCCTTC	360		

Db 9834 ATTCCTAAGGCCCTCTACACCTCTACTCCATCTATGCAATGTTGCCGAGCGCATCCATC 9893  
 Oy 361 GTAAAGTTCACTGACGAGAAACTTGTATAGACTTAAAGAGATATTAATA 420  
 Db 9894 AGCTTCCCGGTTTGTAGAGATACAGGGTTGA-----TCCGAGCTCATC 9938  
 Oy 421 GAGAAAGAAAACCCGTTCTCGGGTACTTCTTCCCAACCAACCCACCGGAAACCTC 480  
 Db 9939 GCCGAGAAAGCCGAACTGTTTCTGCTCTGCCATCCCAACCAACCGGACGACG 9998  
 Oy 481 TTTTCCAGGGGAAAGATTGAGAGATTAAGAAAGGGGTTTCTGTGTATAGACGAA 540  
 Db 9999 CTGACCAAGAAATGTTCAAGAGGTGCTAGAGTGGGAGTACGTTGCTCTGATGAG 10058  
 Oy 541 GCTTACTATCATTTACTCGGAGAAACCTTTCTGAGAACGCCCTCAAAAGGAAAGTACG 600  
 Db 10059 GCTTATGTGAGATTGACGATGATTCAAAGATGTAGATGTTAGGGATTAACACACCTG 10118  
 Oy 601 GTAGTTTGTAGACACTTTCAAA---ATCGGTATGGGAGTTTAAGGTTAGGATTTTA 657  
 Db 10119 ATTTGCTGAGAACTTTTCCAGTTTTCGGCTTTCAGAGATGAGGTTGTTATCT 10178  
 Oy 658 ATAGGAAAGGGGAAATGCTCTCAGAAATTAACAAGTGTAGACTCCCTTCAACGTGAC 717  
 Db 10179 GTCTGACATGGCGAGATTTGCTGAGCAATGTAGAGTAAAGCTCCTTGTGTATAGC 10238  
 Oy 718 TACCCCTCTCAGAGTATGGCAAAAGTTCTCTCAGCGAGGAGAAAGATTTCTATAGAA 777  
 Db 10239 TATCTGCGAGTTGAACCGCAATTTG-CCGCTTGAAGTCTTTGATCTACAAAAGT 10297  
 Oy 778 AAGTATACAGAGGTTGTATACAGAGGAGAAAGATGTACAGCAATATAGAAATAGAA 837  
 Db 10298 AAGGACACAGAA--TGCTCAGGAGAGAGAGGCTTATTGAAACTGAAAGAAATTGAG 10355  
 Oy 838 GGAAGTTGAGGTTTTTCCGATAGGCTAACTTCTGCTTTTCAAGACCTTACCCTCCG 897  
 Db 10356 TGGCTTGAAGTTTACCCGTCAGATGCCAACTTCTGCTTTGAAGCGCAAGAGAGGCG 10415  
 Oy 898 CACGAGGTTTATCAGAGTACTGAAAGGAGATGCTCTGTCGACAGAGTATCTTACATG 957  
 Db 10416 GTGTAG-----AGAGCTGCGCCGAGAGAGGGTTTATTGTAGGAGATGCAAGTTATG 10469  
 Oy 958 GAAGACTCCAAAGTGCCTCAGGGTACGCTAGGCAACCGGAAACCAACACAAGTTT 1017  
 Db 10470 GGGCTTGAAGGCTGTATATTGCAATTACATGCGGTAACGAGAGATATACAGGCTG 10529  
 Oy 1018 CTGGAAGCACTGAGAGAGTATTAATCCCTTTCAAGCTCTCT 1061  
 Db 10530 ATTGAGGCCCTTAAGAAATTTAGAAAGAGTCTTACGCTTTAT 10573

RESULT 11  
 AL646072/c 188050 bp DNA linear BCT 07-DEC-2001  
 LOCUS Ralstonia solanacearum GM11000 chromosome, complete sequence.  
 DEFINITION segment 16/19.  
 ACCESSION AL646072 AL646052  
 VERSION AL646072.1 GI:17429824  
 KEYWORDS Ralstonia solanacearum.  
 SOURCE Ralstonia solanacearum.  
 ORGANISM Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.

REFERENCE  
 1 (bases 1 to 188050)  
 Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Brottier P., Camus J.C., Catolico L., Chandler M., Choisme N., Claudel-Renard C., Cunneac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Signier P., Thebaud P., Whalen M., Wincker P., Levy M., Weisenbach J. and Boucher C.A.  
 Genome sequence of the plant pathogen Ralstonia solanacearum  
 JOURNAL Unpublished

REFERENCE  
 2 (bases 1 to 188050)  
 Boucher C.A.  
 Direct Submission  
 Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
 Cremlieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
 Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
 BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
 Dausset-CEPH, 27 rue Juliette Doda, 75010 Paris, France, LMGM CNRS  
 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
 URGV, 2 rue Gaston Cremlieux, CP5706, 91057 Evry Cedex, France,  
 Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
 F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire  
 INRA, BP27, F31326 Castanet-Tolosan Cedex  
 Christian.Boucher@toulouse.inra.fr  
 http://sequence.toulouse.inra.fr/R.solanacearum.html.

COMMENT  
 location/Qualifiers  
 1. 188050  
 /organism="Ralstonia solanacearum"  
 /strain="GM11000"  
 /db\_xref="taxon:305"  
 73.. 909  
 /gene="RSC2802"  
 /note="R800303"  
 73.. 909  
 /gene="RSC2802"  
 /note="R800303"  
 73.. 909  
 /gene="RSC2802"  
 /function="miscellaneous: hypothetical/global homology"  
 /note="Product confidence : probable  
 Gene name confidence : hypothetical  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="PROBABLE TRANSMEMBRANE PROTEIN"  
 /protein\_id="CAD16509.1"  
 /db\_xref="GI:17429825"  
 /translation="MALLGSMAVGVGNSFAKTLPPALGAGTYRTTIGAILLA  
 LMRPRLRLHRDAGRITALGYTLASMLFLTLSTRLPDITGATIAIEFGPGLYVAL  
 SRRALDFWIGLAVAGLLITVGGRVGIDPLGAVATGAGVCWALTYVGRKVT  
 PAQATSLGMAVALFAIPFGVVOAGPALASLIVAGIGLSSAVPYSLKVALR  
 HSGRTFVSLLEPAIGALAGAVVLEHLRSRWVAIVAIITIASGCAATARSRAA  
 EA"  
 complement(936..2024)  
 /gene="RSC2803"  
 /note="R800302"  
 complement(936..2024)  
 /gene="RSC2803"  
 /function="miscellaneous: hypothetical/partial homology"  
 /note="Product confidence : hypothetical  
 Gene name confidence : hypothetical  
 predicted by Codon\_usage  
 predicted by Framed"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="HYPOTHETICAL PROTEIN"  
 /protein\_id="CAD16510.1"  
 /db\_xref="GI:17429826"  
 /translation="MALETESFPAQTYDQARARELDAARARGLATIRAVHPALPAG  
 EPLSIDPACFLPDRGALLVLTSGIHGVEFCGCGOVLGRDALLFARLAVAGVAL  
 LVHAVNPGFAHLRYVEDNVDLNRNSADPAVAASNPAYIEVDPLLPDTPPPDA  
 DALOOLYLAHMGFAHLRDMNTIGOVAILPDMPFVGSGATCGSTAOACALISRAADAPR  
 LAWIDLTGICGAGHGKRTSGADPRELEPRAITGTWCADYRLIAAGSVSYEGALVD  
 RAGLLEPAVERIVTTLLEFGLTERPAVQAALRADHWLHRHPTSPDDAAIRIGLRARF  
 YCDPPAWKGVVYQARAVYQAAARFSG"  
 complement(2124..3326)  
 /gene="RSC2804"  
 /note="RSC2804: R800301"  
 complement(2124..3326)  
 /gene="RSC2804"  
 /note="RSC2804: R800301"  
 /EC\_number="1.17.4.1"  
 /function="small molecule metabolism; central intermediary

metabolism; 2-prime-deoxyribonucleotide metabolism"  
 /note="Product confidence : putative  
 Gene name confidence : putative  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11  
 /product="PUTATIVE TRANSMEMBRANE RIBONUCLEOSIDE REDUCTASE  
 (SMALL CHAIN) OXIDOREDUCTASE PROTEIN"  
 /protein\_id="CAD16511.1"  
 /db\_xref="GI:17429827"  
 /translation="MLSDDDVKKPAQAQPAQAPVQPPQLQSTPDGGALPPSAT  
 QAAGILGNPNNAQAQSSRRVNAADKRINGATDVNOLYPFYKRWKXYLACANHW  
 MPEIINNSRDIALKDPNGLTEDERLIKRNIGFFVTADSLAANNIVIGYLTQTAPE  
 CROYLLRQAFEEAIHHAQYIVESLGLNEALFNAYHEVQSIRKDEFLIPFIDTLT  
 DPFKGTENQDKLLSLIVFACIMEGLFFVYVGETQILAMGRQNMKGAAEQYQYL  
 LNSPCHNFQIDILQIKLENPHLTAFAEKITELEKAVDLEYRYADTTPRGVLG  
 LNAPFKGYLRFICNRRCQIGLDALFPNEDNPFPMWSEMDLKKERNFFETRVIEQ  
 TCGALSWD"  
 complement(3477..6440)  
 /gene="nrda"  
 /note="Rsc2805; RS00300"  
 complement(3477..6440)  
 /gene="nrda"  
 /EC\_number="1.17.4.1"  
 /function="small molecule metabolism; central intermediary  
 metabolism; 2-prime-deoxyribonucleotide metabolism"  
 /note="Product confidence : putative  
 Gene name confidence : putative  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11  
 /product="PUTATIVE RIBONUCLEOSIDE REDUCTASE 1 (LARGE  
 CHAIN) OXIDOREDUCTASE PROTEIN"  
 /protein\_id="CAD16512.1"  
 /db\_xref="GI:17429828"  
 /translation="NOTAQTEHSCGTANTPAESQPAECTAVASGAGYADYKIIRNGA  
 VVGFPSKIAVMTKFLAVNGGQASARVELVEGUTQNVVRLALRSRPNQGFHFI  
 EDVQVQELMRSGEHDVARAYVLYRERQERQAQETQOQAFIGNVTDGQITR  
 PDMAALNIVSACAGSEDPPEILKETVKNLYEGVPMQSVYDLSAIAARTLIEK  
 DPAYSQARTLILHTIRREILCEEVTAEMAGRYADYFPQIKRGINAGLDDTLAQF  
 DLAKGALDASRLQFNVLQTLDYDRYFLHISDKRIEMPQAFEMRVAMGLSLNEID  
 REARAIEYQLLSDFDMSPTLFNSCTCRSOLSSCYLTIVSDDLGLIYALKENAL  
 LSKFAGGLGNQNTNVRALGSHIKGTNGKSQGVVFLKVYNDTAVAVNOGGKAVCA  
 YLEWHLDIEFLELRKNTGDRRTHDMTANWIPDLFMKRVMEGGEWTLFSPSTCP  
 DLHDKVGRFAEAYLYGYEDKVARGEIKLFKMPALQMLRMLFETGHPWITFKDP  
 CNIRSPQOHVGVHSSNLCTETITNDSETAVCNLGVNLVAHVVKQADGSYALDHD  
 KIKTVRTAMRLDNV IDINTYAVKKARDSNLRHPVGLGIMGFQDALHVLRIPYAND  
 AAVQFADTSMQAVCYAWASTELAEGRYSTYKGSILWDRGVLPODSLKLAERG  
 YLEADLSMTMDGRLRQIOGHMRNSCVAIATPNTISNIGVSACEPTFYQNLVYK  
 SNLSEFTVNDYLVRLDKARGLWDENVADLKYFDGSLRIDRIPDLRLYATATFE  
 VEPQWLVEASRRQRKWDQAOQLNIYMAGASGRMLDPTYKLAWLRLKTYLYLTIGA  
 THVKSTVSRGTLNVAVSGSDAGSVAAGRAPAPASALDTPVAATAPMDAEGAVCT  
 MRPDGPFEECEACQ"  
 complement(6877..7479)  
 /gene="ampb"  
 /note="Rsc2806; RS00299"  
 complement(6877..7479)  
 /gene="ampb"  
 /function="cell processes; protection responses;  
 detoxification"  
 /note="Product confidence : putative  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11

/product="PUTATIVE AMIDASE PROTEIN"  
 /protein\_id="CAD16513.1"  
 /db\_xref="GI:17429829"  
 /translation="MPEVAGPVIRIGADSWADGVCRHPSNVVDAREGLPIDLIVLHNI  
 SLPPARFGTDDVLDFFANTLDCGGAHYFEQGRVRSVSAHFIRRTGACVQFAPCDARA  
 WHAGVSDFRGRKNDFSIGIEIGTDDPPTPAYATAALVRAICAAYPIRALAGH  
 SDIAPGRKTDGPCDFDWAHLRLAKLEAALFPYQAGAAPA"  
 complement(7472..9184)  
 /gene="pehr"  
 /note="Rsc2807; RS00298"  
 complement(7472..9184)  
 /gene="pehr"  
 /function="small molecule metabolism; global functions;  
 global regulatory functions"  
 /note="Product confidence : probable  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11  
 /product="PROBABLE RESPONSE REGULATOR TRANSCRIPTION  
 REGULATOR PROTEIN"  
 /protein\_id="CAD16514.1"  
 /db\_xref="GI:17429830"  
 /translation="MSRQPPFRPPMSKAAVVREPILVVDDEADRELELEISLRMGHD  
 VLVAGLAEAREALTRORFALVLMRGLGDLVROLSATADRTPVAVITAYGSA  
 ENAVELKAGAFDIYAKPLDLDRLSLNALGQRQDPDPSADLAERTSALLPGHS  
 AAMQVRRSLRLARSAPVVISGESGKERAARAIHALSSRSARPVAVNCGAIE  
 NLMEAEFFGVYKGAFTAGTADSDROGFFQAAHGGTLMDELVDPLTMQVKLLRQERR  
 VKTGESREDPDVVRVCASHONLARLYAAGRFREDLFYRLNVLRLMPLREREDV  
 FVLAGVLLEQLAARYGLRPLKRLTRQALQICATFPFGNVELDNLLERAYFAEGES  
 IDVHLGAUGTIDIBRSFLHRAEHGGHPVPAHLPPVAPGHPAHGPHGPHVAHPL  
 GVPQVGGWPDAAIYIPVPMGLMPVVPVPEPAPVPEAPAMPAPVSLPVDLPA  
 YLESVERGVILAAALQTFNRNTASKLLGLSFRLRYRMOQLDIRDPRDIEAAGNGG

Query Match 7.6%; Score 80.6; DB 1; Length 188050;  
 Best Local Similarity 46.2%; Pred. No. 2.3e-11;  
 Matches 475; Conservative 0; Mismatches 524; Indels 30; Gaps 5;

QY 26 TTGAAGCTTACAAAGACGGAGGTCACTCCCGCTCCGTCAGCGCTTCTCTTAACGAATCC 85  
 Db 168373 TGGGCGGTACACGTCGCCGACGCTCCCGCATGGTCAACGTCGACGCGATGAGAAC 168314  
 QY 86 CTTAGACATTTCCGAGAGATATAAACAAGGGCTTAGAAGAATTAATAAAGGTTCCTC 145  
 Db 168313 CTTACCATCTGCCGGCCATGCTGCGCGACGCTCGGCAAGCGCTGCGCGAGTGGCGC 168254  
 QY 146 TCAACAAATACCCAGACCCCAAGCAAGAGTTAAAGCGGTTCTTTCGGGATTTTTCG 205  
 Db 168253 TGAACCGCTACCCGTTGCCACCCGCGCGCTCAAGGCCACGCTCAAGCGCGTGATGC 168194  
 QY 206 CGGTTAAG---GAAGAAATTTAGTTCTTCGTAACGGTTCCGACGAACATCATATAC 262  
 Db 168193 CGGTGCGCGCGCGCGCGACGTGCTGTGGCAACGGCTCGGACGAGATCATCAGCCTGA 168134  
 QY 263 TCTCAATAGCTATAGTGAATTTTACATACCCGTTTACATACCTGTTCCCACTTTCCCA 322  
 Db 168133 TCGTATCGCGCGCGCAAGCCCGCGCGCGCTGCTGGCGCGGCTGCGCGGCTTCGTGA 168074  
 QY 323 TGTACGAGATAGTTCGAAAGTTCTCGGAGACCCCTCGTAAAGGTTCAACTGACGAAA 382  
 Db 168073 TGTATCGATTCGCGCGCAGCTGATGGGCTGAAGTTCTGGGGCGTCCCGCTAGGGCGC 168014  
 QY 383 ACTTTGATATAGCTTAGAAGAAGTATTGAATTAATAGAGAAAGAAAAACCGTTCTTCG 442  
 Db 168013 ATCTCAACCTCGACCGCAAGCGATGCTGCGCGCATGCTGAGCACACCGCGCGTCA 167954  
 QY 443 GGTACTTTGCTTACCCCAACACCCACGGAACCTCTTTTCCAGGGGAAAGATTGAGG 502  
 Db 167953 TCTACCTGGCTATCCGAACAACCCGCGCAACCTGTTTCGACGCGCGCGCATGGATG 167894  
 QY 503 AGATAAGAAACAGGGGTGTTT-----TCTGTGTAATAGACGAAGCCT 544







```

QY 1016 TTCTGGAACCACTGCAG 1032
LOCUS
Db 1252 TTCTAGCCGCTACTGGAG 1268

RESULT 13
AE012885/C
LOCUS
DEFINITION Chlorobium tepidum TLS section 106 of 194 of the complete genome.
ACCESSION AE012885 AE006470
VERSION AE012885.1 GI:21647218
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 11265)
Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,
Dodson, R.J., Deboy, R.K., Gwinn, M.L., Nelson, W.C., Haft, D.H.,
Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,
Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,
Niernman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,
Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,
Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.
The complete genome sequence of Chlorobium tepidum TLS, a
photosynthetic, anaerobic, green-sulfur bacterium
Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)
12093901
2 (bases 1 to 11265)
Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,
Dodson, R.J., Deboy, R.K., Gwinn, M.L., Nelson, W.C., Haft, D.H.,
Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,
Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,
Niernman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,
Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,
Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.
Direct Submission
Submitted (30-APR-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
unpublished
FEATURES
source
Location/Qualifiers
1..11265
/organism="Chlorobium tepidum TLS"
/strain="TLS"
/db_xref="taxon:194439"
complement(114..569)
/gene="mscL"
/note="CT1243"
complement(114..569)
/gene="mscL"
/note="identified by match to PFAM protein family HMM
PF01741"
/codon_start=1
/transl_table=11
/product="large conductance mechanosensitive channel"
/protein_id="AA072473.1"
/db_xref="GI:21647219"
/translation="MLKEFRFALKGNVVDMAVGIIIGAGFALNSLVNLLMPPLG
LLKGVDFSNLFVLKEGTPPGPIYALADAKTAGAVTLNGLFVNALIGFLIMAFVE
LLVRSINRLRSLSKESAAPAVAPQTKCPCFCSIIPKAVRCPNCTSQL"
complement(716..856)
/gene="CT1244"
complement(716..856)
/gene="CT1244"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA072474.1"
/db_xref="GI:21647220"
/translation="MLHRLMTADVWVSWKLPDVSNAGPEFGGDIIVRFYKVGVPVA
FG"

gene
CDS

848..2350
/gene="CT1245"
848..2350
/gene="CT1245"
/note="identified by match to PFAM protein family HMM
PF02754"
/codon_start=1
/transl_table=11
/product="heterodisulfide reductase, subunit C/succinate
dehydrogenase, subunit C"
/protein_id="AA072475.1"
/db_xref="GI:21647221"
/translation="MEHTIRTDLSLRQLQEQTTGNHYACCYQCQCKTAGCPAGFGMDNP
PARIMRLVQAVVEAMQSDSLWYICGCMTCCTARCPQNMETAGTMDVAREMALKSGII
SSDKAKLVTAFHTSFLNTRVSKGRLOELALVNSYKLRTRTFLODAGAGLKMKGKGI
NPTVTAITAKETVEQKQIEKIPKIAEQESHATAAPKPKVKTFRKANEPVRIPTGMITG
YYPGSLSTAKDYDLSIRKMKCDRLRIHEHETEDWNCSSASAHATNKLKSLVLLPARN
QALAQGGFDYVLAFCACINQVTRAKALMESEELNELKSIIMRGDTECEAAQITGIM
QVLEGMDEEIKKRVTHPLKMDPLACYGCLLRPFDAMGYDDPNTKMEAIIIEALG
AKPVDWAYKECCGAGLTMAOQEMIETLTHKAKNATVNGAGGVVAGCPLCHTNLDMR
QEGMKRKYNDGEMPVYVYISDLVAMAGSPAEVALDKHFVPATGMVRNH"
2643..4649
/gene="hdrA-2"
/note="CT1246"
2643..4649
/gene="hdrA-2"
/note="identified by match to PFAM protein family HMM
PF00037"
/codon_start=1
/transl_table=11
/product="heterodisulfide reductase, subunit A"
/protein_id="AA072476.1"
/db_xref="GI:21647222"
/translation="NPEHISTQGFEMAKIGVFVCHCGENIKSKYDTGNLVKALSDHPG
VELICNEYFYCSDPCQEVYKRAIRENNLTGVVAACSRPMHETFRKACAGLNPYM
LELANIREQCSVWVHSDKQATKKAIEITRSVRKRNELKPISVITRRALVIGGG
TAGIQALDIAGAREVILVEREPSIGHMSQLSETFTLDCSQCLTPRVAEAIQHP
NITLVTSYSEVEEDGYIGNFRVYKVKARYVDIDKCTCGDCIQCKPKVPSPEFCG
LGNRTAIYTPFAQVNPVVIDKNRCTYFKNGKICKQCTQICQIENCIDFEMQDTFEEY
EVGAIWATGFIQDTSYGVGYKVDITGLSFERLASASGPTAGKILRPSDGKE
PETVVFIOAGSRDPSKGVKCYCKICCMYAKHAMLYAHKHGHSNKIPFYMDIIRAAG
GYDETRRAIEDEREYLRGRVSKVFEENGLIVRGVDTLLGKVPVEAADMVLAATAI
VPQDAKFAKRGIGYDEYGFYNEAHLKLRPVETATAGIYLCGACOSKPDIPDSVAQ
ASAAAKVYGLFNRQLERPVVAAGSTCAGCGVYACYPYNAIEQKEIRDRNGNL
IKEYASVNPGLCQCGCTVCFRNSIDLAFTEKQIFAEVMAI"
4707..5156
/gene="mvhD"
/note="CT1247"
4707..5156
/gene="mvhD"
/note="identified by match to PFAM protein family HMM
PF02662"
/codon_start=1
/transl_table=11
/product="hydrogenase, methyl-violgen-reducing type, delta
subunit"
/protein_id="AA072477.1"
/db_xref="GI:21647223"
/translation="MSPEFEPKIVAFVCTYCTYAGADIAGTSRLNAPNVRIVRLPCT
GRISPMILKALQKAGSVLYSGCHPGCHTCTAGNYHARRWTVFRALLSFTGPIPER
TRFSWISAEGAKFAELINEITDDTRKGLDPPPTQYQLOKVIETOSTY"
5160..5990
/gene="CT1248"
5160..5990
/gene="CT1248"
/codon_start=1
/transl_table=11
/product="hydrogenase, iron-sulfur binding protein,
putative"
/protein_id="AA072478.1"
/db_xref="GI:21647224"
/translation="MGIOEQYRSEAAALLSGGEVKLVIGFSAGSTADRRRPPFFARTPE
EAEKFVLDAAACIANLSTYLAEGLLSDGKGVGVFLKLDGIRSVNILLISEAQLKEQVW

```



Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsals.  
1 (bases 1 to 1285)

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the sequencing and annotation of the RAEFL cDNAs: Yamada, K., Liu, S.-X., Pham, P. K., Banh, J., Banno, F., Dale, J. M., Goldsmith, A. D., Jiang, P. X., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J., Shinn, P., Southwick, A., Tracy, S. E., Davis, R. W., Ecker, J. R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC) contributed equally to this work as PIs.

# FEATURES

## Location/Qualifiers

1..1570

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

/chromosome="5"

/clone="RAFL08-15-J07 (R11135)"

/note="This clone is in a modified pbluescript vector

(FLC-1) as a BamHI/XhoI insert.

ecotype: Columbia"

1..1570

/gene="At5g10330"

1..133

/gene="At5g10330"

134..1387

/gene="At5g10330"

/codon\_start=1

/evidence="experimental

/product="putative histidinol-phosphate aminotransferase"

/protein\_id="RAK92767.1"

/db\_xref="GI:15292797"

/translation="MGVINVQGSFSSIHSESNLRKSRALKKFKFCIRNRYCAQSS  
SAAVDSESNITMGDSFIRPHLRQAAQPIPLFEVLQAQGRKPEDIVKLDANENPYG  
PPPEYFALGNMKFPYVPDQPSRLRLADQSDSEYILVCGGADELIDLMRCV  
LDPGKIIDCPPTFSMVFDAVNGAGVIGVPRNPDSFSLNVDRIAEVVELEKPCIFL  
TSPNPDGSIISSEDLKILVLEDEAYIEFSGVESRMKWKVKNLYLVRTEF  
KRAGLAGRVGYGAPPLSIEIWLMPAKOPYNVSVAGEVAALAAALNSKGLDVRDALV  
RERELFGLLEVPFLNPYSYFNILCEVTSMDAKKLKEDLAKMGVWVRHNSQEL  
KGYRVSAQKPEHTDVLMECLKQFY"

1388..1570

/gene="At5g10330"

misc\_difference 1553

/gene="At5g10330"

/note="compared to genomic sequence"

/replace="g"

BASE COUNT 444 a 298 c 344 g 484 t

## ORIGIN

Query Match 6.0%; Score 64.2; DB 8; Length 1570;

Best Local Similarity 48.3%; Pred. No. 6.4e-07; Mismatches 193; Indels 0; Gaps 0;

Matches 180; Conservative 0;

QY 190 CTTGCGGATTTTCGGCGTTTAAAGGAAGAAATTTAGTTCTCGGTAAACGGTTTCGGACGAA 249

DB 527 CTGCTCAGACTCTGGTTGGAGTCTGAATACATCTTGTAGGCTGTGGTCTGACGAA 586

QY 250 CTGATATACCTCTCAATAGTATAGGTGAACCTTACATACCCGTTTACATACCTGTT 309

DB 587 CTTATTGATTGATCATGATGCGTGTGTAGATCCCTGGGAGAAGATTATAGATTGTCT 646

QY 310 CCACACCTTTCCCATGTACGAGATAAGTGGCAAGTTCTCGGAAGACCCCTCGTAAAGTT 369

DB 647 CCAACTTCTCAATGATGTGTTGATGGCGTGTGAATGGACGAGTGTCAATAAGTT 706

QY 370 CAACCTGGACGAAACCTTTGATATAGCTTAGAAGAAGTATTGAATTAATAGAGAAGAA 429

DB 707 CCAGAAACCCGATTTCACGTTGAATGACCGCATCGCTGAAGTGTGTAAGAACTAGAA 766

QY 430 AAACCCGTTCTCGGGTACTTTGCTTACCCAAACAACCCACGGGAACCTCTTTCCAGG 489

DB 767 AAACCCCAATGCATATTCTCTAACTTCTCCACAATCCAGATGGGAGTATCATCAGCGAG 826

QY 490 GGAAGATTGAGGAGATAGAAACAGGGGTGTTTCTGTGTATAGACGACGACCTACTAT 549

DB 827 GACGATCTGTTGAAGATTCTAGAAATGCCAATACTTGTGTTCTAGATGAAGCTTACATC 886

QY 550 CATTAATCTCCGGAG 562

DB 887 GAATCTCAGAG 899

Search completed: April 15, 2003, 15:15:13  
Job time : 3628 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:04:46 ; Search time 302 seconds  
(without alignments)  
7941.650 Million cell updates/sec

Title: US-09-905-173-23  
Perfect score: 1065  
Sequence: 1 atgataccaccagaggattaa.....ccctttcaagctctctttaa 1065

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1065	100.0	1065	18 AAT78778	Aquifex histidinol
2	90.8	8.5	13732	21 AAB81484	N. meningitidis pa
3	90.8	8.5	34980	21 AAF21611	Neisseria meningit
4	90.8	8.5	837096	21 AAB81489	N. meningitidis pa
5	60.2	5.7	1664976	19 AAV21209	Methanococcus jann
6	46.2	4.3	1085	24 ABK73007	Bacillus lichenifo
7	42.2	4.0	1092	18 AAT78780	Ammonifex histidin
8	37.6	3.5	588	22 ABA59603	Human foetal liver
9	37.6	3.5	588	22 ABA28180	Probe #6646 for ge

10	37.6	3.5	588	22	AAK07876	Human brain expres
11	37.6	3.5	588	22	AAK33743	Human bone marrow
12	37.6	3.5	588	22	AAI16517	Probe #6450 for ge
13	37.6	3.5	588	22	AAI39468	Probe #8154 used t
14	37.6	3.5	588	24	ABS08616	Human genome-deriv
15	36.8	3.5	9835	23	ABL21524	Drosophila melanog
16	36.4	3.4	6478	22	AAS43416	Chemically pretrea
17	36.4	3.4	6478	24	ABN80200	Human chemically m
18	36.4	3.4	6626	22	AAK46809	DNA transcription
19	36.4	3.4	6804	18	AAV74676	Tumour suppressor
20	36.2	3.4	560	24	ABK77425	Staphylococcus aur
21	35.8	3.4	1098	22	AAH77263	Bacillus clausii g
22	35.8	3.4	1221	22	AAF71952	C glutamicum codin
23	35.8	3.4	10732	21	AAAF1998	Corynebacterium gl
24	35.8	3.4	10732	21	AAAF10594	Corynebacterium gl
25	35.8	3.4	34980	22	AAH68531	Gene encoding a su
26	35.4	3.3	4590	22	AAH24065	C glutamicum codin
27	35.4	3.3	1104	24	ABO86187	Yeast AOD9604-asso
28	35.2	3.3	1350	24	ABD31862	Lactobacillus rham
29	35.2	3.3	5621	24	ABL32205	Lactobacillus rham
30	35	3.3	24187	23	AAAS9529	Human immune syste
31	35	3.3	647	24	ABQ68411	Propionibacterium
32	34.8	3.3	4213	24	ABQ70936	Listeria monocytog
33	34.8	3.3	6631	22	ABA19734	Listeria monocytog
34	34.8	3.3	6631	22	AAK84213	Human nervous syst
35	34.6	3.2	263	23	ABV61754	Human immune/haema
36	34.6	3.2	1341	22	AAI93087	Human prostate exp
37	34.6	3.2	2483	24	ABK84870	Human polynucleoti
38	34.6	3.2	8539	24	ABK84868	Nematode infection
39	34.6	3.2	19243	22	AAK71662	Nematode resistanc
40	34.6	3.2	375	22	AAK06304	Human immune/haema
41	34.4	3.2	375	22	AAK31971	Human brain expres
42	34.4	3.2	375	22	ABS06734	Human bone marrow
43	34.4	3.2	12225	18	AAT89257	Human genome-deriv
44	34.4	3.2	12616	18	AAT89255	Human bg gene (sho
45	34.4	3.2	12616	18	AAT89255	Human bg gene (lon

## ALIGNMENTS

RESULT 1  
AAT78778  
ID AAT78778 standard; DNA; 1065 BP.  
XX  
AC AAT78778;  
XX  
DT 18-FEB-1998 (first entry)  
XX  
DE Aquifex histidinol-phosphate aminotransferase VF5/HPA DNA.  
XX  
KW Histidinol-phosphate aminotransferase; VF5/HPA; chiral compound;  
KW ss.  
XX  
OS Aquifex sp. strain VF5.  
XX  
PN WO9729187-A1.  
XX  
PD 14-AUG-1997.  
XX  
PF 21-JAN-1997; 97WO-US01094.  
XX  
PR 08-MAY-1996; 96US-0646590.  
PR 09-FEB-1996; 96US-0599171.  
XX  
PR (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
PI Swanson RV, Warren PV;  
XX  
DR WPI; 1997-415343/38.  
DR P-PSDE; AAW24254.  
XX  
PT New transaminase(s) and aminotransferase(s) derived from host cells

PT - used for producing enzymes, and hybridisation probes for a cDNA or genomic library

XX Claim 1; Fig 7; 95pp; English.

XX This genomic DNA encodes histidinol-phosphate aminotransferase  
CC VF5/HPA (see AW24254) of Aquifex VF5, a strictly  
CC chemolithoautotrophic, marine eubacterium which grows optimally at  
CC 85-90 deg C and pH 6.8 in high salt medium. The VF5/HPA sequence  
CC can be amplified from a pBluescript vector that contains the DNA by  
CC PCR (see AAF78794-95). The DNA can be used for recombinant  
CC production of the aminotransferase and to develop hybridisation  
CC probes. Claimed thermostable transaminases and aminotransferases  
CC (AAW24248-57) can be produced from native or recombinant host cells  
CC for use with L- and/or D-amino acids for production of optically  
CC pure chiral compounds used in the pharmaceutical, agricultural and  
CC other industries. A method is claimed for transferring an amino  
CC group from an amino acid to an alpha-keto acid using a claimed  
CC enzyme.

XX Sequence 1065 BP; 344 A; 220 C; 252 G; 249 T; 0 other;

Query Match 100.0%; Score 1065; DB 18; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAGGAACCTTGAAGCTTACAAAGACGAGGTCACTCCCGCTCC 60  
DB 1 ATGATACCCAGAGGATTAAGGAACCTTGAAGCTTACAAAGACGAGGTCACTCCCGCTCC 60  
QY 61 GTCAGGCTTCTCTACGAAATTCCTTACGACTTTCCCGAGGAGATATAAACAAGGGCC 120  
DB 61 GTCAGGCTTCTCTACGAAATTCCTTACGACTTTCCCGAGGAGATATAAACAAGGGCC 120  
QY 121 TTAGAAGAAATTAAGGCTTCCCTTCAACAATAATACCCAGACCCCGAAGGAAAGAGTTA 180  
DB 121 TTAGAAGAAATTAAGGCTTCCCTTCAACAATAATACCCAGACCCCGAAGGAAAGAGTTA 180  
QY 181 AAACGGGTTCTCGGGATTTTTCGGGCTTAAGGAAGAAATTTAGTTCCTCGTAAACGGT 240  
DB 181 AAACGGGTTCTCGGGATTTTTCGGGCTTAAGGAAGAAATTTAGTTCCTCGTAAACGGT 240  
QY 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
DB 241 TCGGACGAACTCATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
QY 301 ATACCTGTCCACCTTTCCCATGTACGAGATAAGTCGGAAGTTCTCGAAGACCCCTC 360  
DB 301 ATACCTGTCCACCTTTCCCATGTACGAGATAAGTCGGAAGTTCTCGAAGACCCCTC 360  
QY 361 GTAAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAGTATTGAATTAATA 420  
DB 361 GTAAAGGTTCAACTGGAGAAACCTTTGATATAGACTTAGAAGAAGTATTGAATTAATA 420  
QY 421 GAGAAGAAACCCGTTCTCGGTACTTTTCTTACCCCAACACCCACCGGGAACCTC 480  
DB 421 GAGAAGAAACCCGTTCTCGGTACTTTTCTTACCCCAACACCCACCGGGAACCTC 480  
QY 481 TTTTCCAGGGGAAAGATTGAGGAGATAAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
DB 481 TTTTCCAGGGGAAAGATTGAGGAGATAAGAAACAGGGGTGTTTCTGTGTAATAGACGAA 540  
QY 541 GCCTACTATCATTTACTCCGGAGAACCTTTCTGGAAGACGGCTCAAAAGGGAAGATACG 600  
DB 541 GCCTACTATCATTTACTCCGGAGAACCTTTCTGGAAGACGGCTCAAAAGGGAAGATACG 600  
QY 601 GTAGTTTTCAGACACTTTCAAAATCCGTTATGCGAGCTTTTAAAGGTAGGATTTTAATA 660  
DB 601 GTAGTTTTCAGACACTTTCAAAATCCGTTATGCGAGCTTTTAAAGGTAGGATTTTAATA 660  
QY 661 GGAAGAGGGGAAATCGTCTCGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTTAC 720  
DB 661 GGAAGAGGGGAAATCGTCTCGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTTAC 720

QY 721 CCCTCTCAGGTGATGGCAAAAAGTTCTCTCACGGAGGAGAGAAATTCCTAATGGAAGG 780  
DB 721 CCCTCTCAGGTGATGGCAAAAAGTTCTCTCACGGAGGAGAGAAATTCCTAATGGAAGG 780  
QY 781 ATACAGGAGGTTGTAACAGACGAGAGAAAGATGTACACGAAATGAAGAAATAGAAGGA 840  
DB 781 ATACAGGAGGTTGTAACAGACGAGAGAAAGATGTACACGAAATGAAGAAATAGAAGGA 840  
QY 841 GTTGAGGTTTTCAGGTAAGCTTCTTCTTTCAGAACGCTTACCCGCGCCAC 900  
DB 841 GTTGAGGTTTTCAGGTAAGCTTCTTCTTTCAGAACGCTTACCCGCGCCAC 900  
QY 901 GAGGTTTATCAGGAGCTACTGAAAGGAGTCTCTCGTACGAAACGATCTTACATGAA 960  
DB 901 GAGGTTTATCAGGAGCTACTGAAAGGAGTCTCTCGTACGAAACGATCTTACATGAA 960  
QY 961 GGACTCCAAAAGTGCTCAGGTAAGCGTAGGAAACCGGAGAAACAAACAAGTTCTG 1020  
DB 961 GGACTCCAAAAGTGCTCAGGTAAGCGTAGGAAACCGGAGAAACAAACAAGTTCTG 1020  
QY 1021 GAAGCACTGGAGGAGATATAAATFCCCTTTCAAGCTCTCTTTAA 1065  
DB 1021 GAAGCACTGGAGGAGATATAAATFCCCTTTCAAGCTCTCTTTAA 1065

# RESULT 2

AAA81484

ID AAA81484 standard; DNA; 13732 BP.

XX AC AAA81484;

XX DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_32 SEQ ID NO:32.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; Neisseria; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.

XX OS Neisseria meningitidis.

XX PN WO200022430-A2.

XX PD 20-APR-2000.

XX PF 08-OCT-1999; 99WO-US23573.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlato V;

PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
PT used in the diagnosis and treatment of N. meningitidis infection and  
PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 589-593; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF









Db 209810 TGATAGCTTTACGACCCCTCAGCAAAATCGGTTTTCGCGACTCGGTATCGGTTATCGG 209869  
 Qy 659 TAGGGAAGGGGAAATCTCTCAGAAATTAACAGAGGTGAGACTCCCTTCAACGTGACCT 718  
 Db 209870 CAGGCTCCCGAAGTATCATCGGCACTGCAAAATCTTCGCGCCCTTACAATATGAACC 209929  
 Qy 719 ACCCCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGGAAGAGAATTCCTAATGAAA 778  
 Db 209930 AATTGAGCCTGACCACTGCCAAATCGCCCTGCGGCACTACGCAATATCTCT--GCCA 209986  
 Qy 779 AGATACAGGAGGTCTTACAGAGGAGAAAGGATGTACGACGAAATGAAGAAATAGAAG 838  
 Db 209987 ACATCGACAGCTGAAAGCAAGCAAGCAAGGATGTCGCGCAATGGGCAAAATATGCC 210046  
 Qy 839 GAGTTGAGGTTTTTCCGAGTAAGGCTAACTTCTT 872  
 Db 210047 GTCTGAACACCTTTCAAGTCAGGCAAACTTCAT 210080

## RESULT 5

AAV21209/c  
 ID AAV21209 standard; DNA; 1664976 BP.  
 AC AAV21209;  
 XX  
 XX  
 DT 10-NOV-1998 (first entry)  
 XX  
 XX Methanococcus jannaschii circular chromosome.  
 XX  
 KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 KW genome; autotrophic; extrachromosomal element; identification; ds.  
 OS Methanococcus jannaschii.  
 XX  
 XX WO9807830-A2.  
 XX  
 XX 26-FEB-1998.  
 XX  
 XX 22-AUG-1997; 97WO-US14900.  
 XX  
 XX 22-AUG-1996; 96US-0024428.  
 XX  
 XX (GENO-) INST GENOMIC RES.  
 XX (UNII) UNIV ILLINOIS FOUND.  
 XX (YUJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX  
 XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 XX WPI; 1998-169145/15.  
 XX  
 XX Complete genome sequence of methano-genic archaeon, Methanococcus  
 PT jannaschii - useful in identification of M. jannaschii genome  
 PT fragment

Claim 13; Page 152-585; 614pp; English.

The present sequence represents the complete 1.66-megabase pair genome sequence of the Methanococcus jannaschii circular chromosome. The present invention describes M. jannaschii open reading frames from the genome sequence. The invention also describes a computer based system for identifying fragments of the M. jannaschii genome that are homologous to target nucleotide sequences, comprising: (a) data storage means comprising the nucleotide sequence of the 1664976, 58407 or 16550 bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide sequence at least 99.9% identical to it; (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means to identify a homologous sequence, and (c) retrieval means for obtaining the homologous sequence. The method, which is based on whole genome random sequencing of an autotrophic archaeon M. jannaschii, the genome of which consists of 3 physically distinct elements, a large circular chromosome (the 1664976 bp sequence given in AAV21209), a large circular extra-chromosomal element (the 58407 bp sequence given in AAV21210), and a small circular extra-chromosomal element (the 16550 bp sequence given

CC in AAV21211), can be used in the identification of M. jannaschii genome fragment.

XX  
 SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other; Query Match 5.7%; Score 60.2; DB 19; Length 1664976; Best Local Similarity 47.7%; Pred. No. 2.9e-06; Matches 479; Conservative 0; Mismatches 493; Indels 33; Gaps 9;

Qy 62 TCAGGCTTTCCTCTAACGAATTCCTCAGACTTTCCCGAGGAGATATAACAAAGGCCCT 121  
 Db 586070 TAAACTCTGGTCTTAATGAAATCTCTGGGACCATCTCCAAATAATAAAGAAAAATTT 586011  
 Qy 122 TAGAAGAATTAATAAGGTTCCCTTGACAATAACCCAGACCCGGAAGGAGAGTTAA 181  
 Db 586010 TAGATGAAATTGACAAAATTC-----ACCAATATCCAGAGCCAGTAATCCAATTTAA 585957  
 Qy 182 AAGCGGTTCTCGGATTTTTCGGGTTTAAAGAGAATAATTTA---GTTCTCGGTAAAG 238  
 Db 585956 TGAAGAGTTAAGCAAAATTTTGAATGTTGATGAGGAAACAATAATGTTGGAGGAGATG 585897  
 Qy 239 GTTCGACGAACTCATATACCTCTCAATAGCTATAGTGAACTTTACATACCCGTTT 298  
 Db 585896 GAGCTGATGAGATTATAGACACAATATTTAGAACGTTTGTGATGATGAGATGAGGTTA 585837  
 Qy 299 ACATACCTGTTCCACCTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCC 358  
 Db 585836 TAAATCCAATTCACAATTTACCCAATATAGAGTTTCAGCAACAATCCACAATGCTAATA 585777  
 Qy 359 TCCTAAGGTTCAACTGGACGAAACTTTGATATAGACTTAGAAGAAGATATGTAATTTAA 418  
 Db 585776 TAAATATGCTAAATATGATAGGAGAAAGACTTTTAAATTTGAATTTGAAAGTGCTTTAA 585717  
 Qy 419 TAGA---GAAAGAAAACCCGTTCTCGGTACTTTGCTTACCACCAACCCAGCGGAA 475  
 Db 585716 ATAATATAACAGATAAAACGAAAGTTATTTTCTCTGCACCTCCAAATAATCCACAGAA 585657  
 Qy 476 ACCTCTTTTCCAGGGGAAAGATTGA---GGAGATAAGAAACAGGGGTTTTCTGTGTAA 532  
 Db 585656 ATATAATAAGAAATAGAGATGTAGAGAGGGTTATCAATGAAACAGACGCTTTAGTTGTTA 585597  
 Qy 533 TAGCGAAGCCTACTATCATTTACTCCCGAGAAACCTTT-----CTGGAAGACCGCTCA 586  
 Db 585596 TTAGCATGATACATTCAGTATGCTTAAAGAAATATGATTTGACTCAAAAGCCCTG 585537  
 Qy 587 AAAGGAAGATACGGTAGTTTGGAGACACTTTCAAAATC---GGTATGGCGAGTTTAA 643  
 Db 585536 AATATGATAATGTTATTTTAAAGAACCTTTTCAAGGCTTTTGGTTTAGCAGGAATGA 585477  
 Qy 644 GGTAGGGATTTTAAAGGAGGGGAAATCGTCTCAGAATTAACAAGGTGAGACTCC 703  
 Db 585476 GGGTTGGTTATGGTGTGCAATAAAAAATCATAGATTATATGATGAGAGTTAAGCCAA 585417  
 Qy 704 CCTTCAAGCTACCTACCCCTCTCAGGTGATGCGCAAAAGTTCTCTCAGGAGGGAAGAG 763  
 Db 585416 TATTAGCTTAAAGGTTAAGTCAAGTTTGTGCCATACTGCATTAAAGAT---AGAG 585360  
 Qy 764 AATTCCTAATCGAAGAAATACAGAGGTTGTAAACAGCGGAGAAAGGATGTACAGCGAAA 823  
 Db 585359 AATTCCTTGAAGATGTGTAGAGATGGAATTAAGATAGAGATGCTCTACAATGGAT 585300  
 Qy 824 TGAAGAAAATAGAAGAGTTGAGGTTTTTCCGAGTAAGGCTTAAC---TCTTGGCTTTTCA 880  
 Db 585299 TGAAGAGTTTAAAGATATTAAGGTTTATCTTCAGAGCTTAATATCTATTGGTTGAAT 585240  
 Qy 881 GAACGCTTACCCGCCACAGGTTTATCAGGAGCTACTGAAAGGATGCTCCTCGTCA 940  
 Db 585239 TAAAAACAATGAAACGCAAAAGAAATTTTGTGAGGAACATTTAAAAAGAGGTGTTATGTTA 585180  
 Qy 941 GGAAGTATCTTACATCGAAGG---ACTCCAAAAGTCCCTCAGGGTAAGCGGTAGGGAAC 997  
 Db 585179 GAGATTGCACATCTCTTTGATGGTTTATGGGATAATATTGTTAGAGTATCAATAGGACGT 585120





CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;  
Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739  
Db 309 CAGGCTTAGCAATCTGTGGGACAAACATCAAAATGTGCAACTTATGTGTGTAGTTA 368  
QY 740 AAGTTCTCCTCAGGAGGAGAGAAATTCCTAATGGAAGAGATACAGAGGTTGTAAACAG 799  
Db 369 CAGAAGAGAGATGTGGGGAGGAGGATATTGGAACAAATAATGGCTCAAAACT 428  
QY 800 AGCGAGAAAGGATGTACGACCAATGAAGAAATAGAGAGGTTGAGGTTTTCGAGTA 859  
Db 429 TCCCAATCTGATGAAGACATAAATAAATCAATCAAGAAATTTTAAGGAAGTCCAAGTA 488

RESULT 10  
AAK07876  
ID AAK07876 standard; DNA; 588 BP.

XX AC AAK07876;  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe SEQ ID NO: 7867.  
XX KW Human; brain expressed exon; gene expression analysis; probe;  
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX KW epilepsy; cancer; ss.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PS Example 4; SEQ ID NO: 7867; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is one of the probes of the  
XX CC invention.  
XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739  
Db 309 CAGGCTTAGCAATCTGTGGGACAAACATCAAAATGTGCAACTTATGTGTGTAGTTA 368  
QY 740 AAGTTCTCCTCAGGAGGAGAGAAATTCCTAATGGAAGAGATACAGAGGTTGTAAACAG 799  
Db 369 CAGAAGAGAGATGTGGGGAGGAGGATATTGGAACAAATAATGGCTCAAAACT 428  
QY 800 AGCGAGAAAGGATGTACGACCAATGAAGAAATAGAGAGGTTGAGGTTTTCGAGTA 859  
Db 429 TCCCAATCTGATGAAGACATAAATAAATCAATCAAGAAATTTTAAGGAAGTCCAAGTA 488

RESULT 11  
AAK33743  
ID AAK33743 standard; DNA; 588 BP.

XX AC AAK33743;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8300.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PS Example 4; SEQ ID NO: 8300; 658pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC bone marrow. They can be used to measure gene expression in bone marrow  
XX CC samples, which may enable the improved diagnosis and treatment of cancers  
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX CC the probes of the invention.  
XX SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;  
Best Local Similarity 50.6%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 680 CAGAAATTAACAAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGCGAA 739

```

Db 309 CAGGCCTTAGCAATCTGTGGGACAAACATCAAAATCTGCCAACTTATGTGTTGTGTAGTTA 368
QY 740 AAGTTCTCTCCACGGAGGAGAGAGAAATTCCTAATGGAAAAGATACACGAGGTTGTAAACAG 799
Db 369 CAGAGAAAGAGATGTCGGGGAGGAAAGGATATTGGAAAACAAAATAATGGCTGAAAAC 428
QY 800 AGCGAGAAAGAGATGACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTCCTCCGAGTA 859
Db 429 TCCCAATCTGATGAAAGACATAATATAACATCCCAAGATTTTAAGGAAGTCCCAAGTA 488

RESULT 12
AAI16517
ID AAI16517 standard; DNA; 588 BP.
XX
AC AAI16517;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #6450 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 6450; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;
Best Local Similarity 50.6%; Pred. No. 0.45;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTAACAGGTGACACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGGCAA 739
Db 309 CAGGCCTTAGCAATCTGTGGGACAAACATCAAAATGTGCCAACTTATGTGTTGTGTAGTTA 368
QY 740 AAGTTCTCTCCACGGAGGAGAGAGAAATTCCTAATGGAAAAGATACACGAGGTTGTAAACAG 799
Db 369 CAGAGAAAGAGATGTCGGGGAGGAAAGGATATTGGAAAACAAAATAATGGCTGAAAAC 428
QY 800 AGCGAGAAAGAGATGACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTCCTCCGAGTA 859
Db 429 TCCCAATCTGATGAAAGACATAATATAACATCCCAAGATTTTAAGGAAGTCCCAAGTA 488

```

```

Db 369 CAGAGAAAGAGATGTCGGGGAGGAAAGGATATTGGAAAACAAAATAATGGCTGAAAAC 428
QY 800 AGCGAGAAAGAGATGACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTCCTCCGAGTA 859
Db 429 TCCCAATCTGATGAAAGACATAATATAACATCCCAAGATTTTAAGGAAGTCCCAAGTA 488

RESULT 13
AAI39468
ID AAI39468 standard; DNA; 588 BP.
XX
AC AAI39468;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #8154 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 8154; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 588 BP; 274 A; 71 C; 111 G; 132 T; 0 other;

Query Match 3.5%; Score 37.6; DB 22; Length 588;
Best Local Similarity 50.6%; Pred. No. 0.45;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 680 CAGAAATTAACAGGTGAGACTCCCTTCAACGTGACCTACCCCTCTCAGGTGATGGCAA 739
Db 309 CAGGCCTTAGCAATCTGTGGGACAAACATCAAAATGTGCCAACTTATGTGTTGTGTAGTTA 368
QY 740 AAGTTCTCTCCACGGAGGAGAGAGAAATTCCTAATGGAAAAGATACACGAGGTTGTAAACAG 799
Db 369 CAGAGAAAGAGATGTCGGGGAGGAAAGGATATTGGAAAACAAAATAATGGCTGAAAAC 428
QY 800 AGCGAGAAAGAGATGACGACGAAATGAAGAAATAGAGAGGTTGAGGTTTTCCTCCGAGTA 859
Db 429 TCCCAATCTGATGAAAGACATAATATAACATCCCAAGATTTTAAGGAAGTCCCAAGTA 488

```

```

RESULT 14
ABS08616
ID ABS08616 standard; DNA; 588 BP.
XX
AC ABS08616;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID NO 8607.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-234687P.
XX
XX 27-SEP-2000; 2000US-236359P.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
XX Claim 1; SEQ ID NO 8607; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridise at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX CC probes/open reading frames (ORF). The probes are used for gene
XX expression analysis, and for identifying exons in a gene, particularly
XX

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 14:00:46 ; Search time 63 Seconds  
(without alignments)  
5184.298 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgatacccccagaggatttaa.....cccttcaagctctctttaa 1065

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1065	100.0	1065	1	US-08-599-171A-23		Sequence 23, Appl
2	1065	100.0	1065	2	US-08-646-590B-23		Sequence 23, Appl
3	1065	100.0	1065	3	US-09-069-226-23		Sequence 23, Appl
4	1065	100.0	1065	4	US-09-412-184-23		Sequence 23, Appl
5	61	5.7	7218	1	US-08-232-463-14		Sequence 14, Appl
6	42.2	4.0	1092	2	US-08-646-590B-35		Sequence 35, Appl
7	42.2	4.0	1092	4	US-09-412-184-35		Sequence 35, Appl
8	34.4	3.2	12225	2	US-08-822-445-11		Sequence 11, Appl
9	34.4	3.2	12225	4	US-09-396-540-11		Sequence 11, Appl
10	34.4	3.2	12616	4	US-08-822-445-9		Sequence 9, Appl
11	34.4	3.2	12616	4	US-09-396-540-9		Sequence 9, Appl
12	33.2	3.1	1100	1	US-08-234-939-7		Sequence 7, Appl
13	33.2	3.1	1100	1	US-08-558-865-7		Sequence 7, Appl
14	33.2	3.1	1268	1	US-08-234-939-6		Sequence 6, Appl
15	33.2	3.1	1268	1	US-08-558-865-6		Sequence 6, Appl
16	33.2	3.1	1385	1	US-08-234-939-5		Sequence 5, Appl
17	33.2	3.1	1385	1	US-08-558-865-5		Sequence 5, Appl
18	33.2	3.1	1859	1	US-08-234-939-4		Sequence 4, Appl
19	33.2	3.1	1859	1	US-08-558-865-4		Sequence 4, Appl
20	33.2	3.1	1933	1	US-08-234-939-3		Sequence 3, Appl
21	33.2	3.1	1933	1	US-08-558-865-3		Sequence 3, Appl
22	33.2	3.1	3426	1	US-08-234-939-1		Sequence 1, Appl
23	33.2	3.1	3426	1	US-08-558-865-1		Sequence 1, Appl
24	33.2	3.1	3426	3	US-08-654-025-6		Sequence 1, Appl
25	32.8	3.1	180	4	US-08-894-818B-31		Sequence 31, Appl
26	32.4	3.0	1770	4	US-08-943-731-146		Sequence 146, App
27	32.4	3.0	24183	4	US-08-943-731-3		Sequence 3, Appl

C 28 32.2 3.0 477 3 US-09-026-343-12 Sequence 12, Appl  
C 29 32.2 3.0 477 4 US-09-362-871-12 Sequence 12, Appl  
C 30 31.6 3.0 5134 2 US-08-310-912A-157 Sequence 157, App  
C 31 31.6 3.0 5134 4 US-09-301-085-157 Sequence 157, App  
C 32 31.6 3.0 5134 5 PCT-US95-04589-157 Sequence 157, App  
C 33 31.6 3.0 10968 2 US-08-680-327-2 Sequence 2, Appl  
C 34 31.6 3.0 10968 4 US-09-228-246-1 Sequence 1, Appl  
C 35 31.2 2.9 70000 4 US-09-851-896-3 Sequence 3, Appl  
C 36 31 2.9 289 4 US-09-007-005-17 Sequence 17, Appl  
C 37 31 2.9 289 4 US-09-244-796-17 Sequence 17, Appl  
C 38 30.6 2.9 1236 4 US-08-964-652-1 Sequence 1, Appl  
C 39 30.6 2.9 1423 1 US-07-829-954-1 Sequence 1, Appl  
C 40 30.6 2.9 1423 1 US-07-994-423-1 Sequence 1, Appl  
C 41 30.6 2.9 1423 1 US-08-421-891-1 Sequence 1, Appl  
C 42 30.6 2.9 25002 4 US-08-961-527-48 Sequence 48, Appl  
C 43 30.4 2.9 115 1 US-08-488-470A-11 Sequence 11, Appl  
C 44 30.4 2.9 115 1 US-07-946-239-3 Sequence 3, Appl  
C 45 30.4 2.9 115 1 US-08-484-505A-11 Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-08-599-171A-23  
; Sequence 23, Application US/08599171A  
; Patent No. 5814473  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/599, 171A  
FILING DATE: Concurrently  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HERRON, CHARLES J.  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 331400-38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA

US-08-599-171A-23  
Query Match 100.0%; Score 1065; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 ATGATACCCAGAGGATTAGGAAGCTTGAAGCTTACAGCGAGGTCACTCCCGCTCC 60

```

US-08-646-590B-23
; Sequence 23, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646.590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1065 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1062
;
US-08-646-590B-23
Query Match 100.0%; Score 1065; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAGGAACCTTGAAGCTTACAAGACGGAGGTCACCTCCCGCCTCC 60
QY 61 GTACAGGCTTCTCTAACAAGATTTCCCTACGACTTCCCGAGGAGATTAACAAGAGGCC 120
Db 61 GTACAGGCTTCTCTAACAAGATTTCCCTACGACTTCCCGAGGAGATTAACAAGAGGCC 120
QY 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAATATCCCGAGACCCCGAAGCGAAGAGTTA 180
Db 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAATATCCCGAGACCCCGAAGCGAAGAGTTA 180
QY 181 AAAGCGGTTCTTCGGGATTTTTCGGGCTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
Db 181 AAAGCGGTTCTTCGGGATTTTTCGGGCTTAAGGAAGAAATTTAGTTCTCGGTAACGGT 240
QY 241 TCAGGACCACTCATATACCTACCTCTCAATAGCTATAGTGAACCTTTACATACCGCTTAC 300
Db 241 TCAGGACCACTCATATACCTACCTCTCAATAGCTATAGTGAACCTTTACATACCGCTTAC 300

```



QY 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACTAC 720  
DB 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACTAC 720  
QY 721 CCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGAGAGAAATTCCTAATGAAAG 780  
DB 721 CCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGAGAGAAATTCCTAATGAAAG 780  
QY 781 ATACAGGAGGTGTTAACAGAGGAGAGAGATGTACGACGAAATGAAGAAATAGAGGA 840  
DB 781 ATACAGGAGGTGTTAACAGAGGAGAGAGATGTACGACGAAATGAAGAAATAGAGGA 840  
QY 841 GTTGAAGTTTCCAGTAAGCTAACTTCTGCTTTTCAAGAGCCTTACCCGCCAC 900  
DB 841 GTTGAAGTTTCCAGTAAGCTAACTTCTGCTTTTCAAGAGCCTTACCCGCCAC 900  
QY 901 GAGGTTTATCAGGAGTACTGAAAAGGAGTGTCTGCTGTCAGAGAGTATCTTACATGGA 960  
DB 901 GAGGTTTATCAGGAGTACTGAAAAGGAGTGTCTGCTGTCAGAGAGTATCTTACATGGA 960  
QY 961 GGAAGTCCAAAGTCTCAGGCTAGGCTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 961 GGAAGTCCAAAGTCTCAGGCTAGGCTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 GAAGCACTGGAGGAGATATAAAATCCCTTTTCAAGCTCTCTTTAA 1065  
DB 1021 GAAGCACTGGAGGAGATATAAAATCCCTTTTCAAGCTCTCTTTAA 1065

RESULT 4  
US-09-412-184-23  
Sequence 23, Application US/09412184  
Patent No. 6268188  
GENERAL INFORMATION:  
APPLICANT: Warren, Patrick V.  
APPLICANT: Swanson, Ronald V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/412,184  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,590  
FILING DATE: 08-May-1996  
APPLICATION NUMBER: 08/599,171  
FILING DATE: 09-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01094  
FILING DATE: 21-January-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/017001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1062  
US-09-412-184-23  
Query Match 100.0%; Score 1065; DB 4; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATATCCCCAGAGGATTAAGGAACTTGAAGCTTACAAGACGAGGTCACTCCCGCTCC 60  
DB 1 ATGATATCCCCAGAGGATTAAGGAACTTGAAGCTTACAAGACGAGGTCACTCCCGCTCC 60  
QY 61 GTCAGGCTTTCCTCTACAGAAATTCCTCCCTAGCACTTCCCGAGGAGATAAACAAGAGGCC 120  
DB 61 GTCAGGCTTTCCTCTACAGAAATTCCTCCCTAGCACTTCCCGAGGAGATAAACAAGAGGCC 120  
QY 121 TTAGAAGAAATTAAGGAGTTCCCTTTGAACAAATACCCAGACCCCGAAGGAAAGAGTTA 180  
DB 121 TTAGAAGAAATTAAGGAGTTCCCTTTGAACAAATACCCAGACCCCGAAGGAAAGAGTTA 180  
QY 181 AAGCGGTTCTTGGGAGTTTTCCTGGGCTTAAAGGAGAAATTTAGTTCGCGTAACGGT 240  
DB 181 AAGCGGTTCTTGGGAGTTTTCCTGGGCTTAAAGGAGAAATTTAGTTCGCGTAACGGT 240  
QY 241 TCGACGAACTCATATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
DB 241 TCGACGAACTCATATATACCTCTCAATAGCTATAGTGAACCTTTACATACCCGTTTAC 300  
QY 301 ATACCTGTTCACCTTTCCCATGTACGAGATAGTCGGAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTTCACCTTTCCCATGTACGAGATAGTCGGAAGTTCTCGGAAGACCCCTC 360  
QY 361 GTAAGGTTCAACTGGAGAAACTTTGATATAGACTTAGAAGAACTATTGAATTAATA 420  
DB 361 GTAAGGTTCAACTGGAGAAACTTTGATATAGACTTAGAAGAACTATTGAATTAATA 420  
QY 421 GAGAAAGAAACCCGTTCTCGGCTACTTGTCTTACCCAAACACCCAGGGAAGCTC 480  
DB 421 GAGAAAGAAACCCGTTCTCGGCTACTTGTCTTACCCAAACACCCAGGGAAGCTC 480  
QY 481 TTTCCAGGGGAAAGATTGAGGAGATAGAAACAGGGGTGTTTCTGTGTAATAGAGCAA 540  
DB 481 TTTCCAGGGGAAAGATTGAGGAGATAGAAACAGGGGTGTTTCTGTGTAATAGAGCAA 540  
QY 541 GCCTACTATCATCTCCGAGAAACCTTTCTGGAAGACCGCTTCAAAAGGAGATACG 600  
DB 541 GCCTACTATCATCTCCGAGAAACCTTTCTGGAAGACCGCTTCAAAAGGAGATACG 600  
QY 601 GTAGTTTGTAGGACACTTTCAAAATCGGTATGGGAGTTTAAAGGTAGGATTTAATA 660  
DB 601 GTAGTTTGTAGGACACTTTCAAAATCGGTATGGGAGTTTAAAGGTAGGATTTAATA 660  
QY 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACTAC 720  
DB 661 GGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACGTCACTAC 720  
QY 721 CCCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGAGAGATTCCTAATGAAAG 780  
DB 721 CCCTCTCAGGTGATGGCAAAAGTCTCTCAGGAGGAGAGATTCCTAATGAAAG 780  
QY 781 ATACAGGAGGTGTTAACAGAGGAGAGAGATGTACGACGAAATGAAGAAATAGAGGA 840  
DB 781 ATACAGGAGGTGTTAACAGAGGAGAGAGATGTACGACGAAATGAAGAAATAGAGGA 840  
QY 841 GTTGAAGTTTCCAGTAAGCTAACTTCTGCTTTTCAAGAGCCTTACCCGCCAC 900  
DB 841 GTTGAAGTTTCCAGTAAGCTAACTTCTGCTTTTCAAGAGCCTTACCCGCCAC 900

```
QY 901 GAGTTTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAGCGTATCTTACATGGAA 960
Db 901 GAGTTTATCAGGAGTACTGAAAGGATGCTCTGTCAGGAGCGTATCTTACATGGAA 960
QY 961 GGAATCAAAAGTCCCTCAGGTTAAGCGTAGGGAACCGGAGAAACAAACAAAGTTCTG 1020
Db 961 GGAATCAAAAGTCCCTCAGGTTAAGCGTAGGGAACCGGAGAAACAAACAAAGTTCTG 1020
QY 1021 GAAGCACTGGAGGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065
Db 1021 GAAGCACTGGAGGAGATATAAATCCCTTTCAAGCTCTCTTTAA 1065

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-F1s
US-08-232-463-14

Query Match 5.7%; Score 61; DB 1; Length 7218;
Best Local Similarity 1.9%; Pred. No. 2.4e-09;
Matches 7; Conservative 229; Mismatches 139; Indels 0; Gaps 0;

QY 473 GAAACCTCTTTCCAGGGAAGATTGAGGATAGAAACAGAGGGGCTTTCTGTGTAA 532
Db 1448 GAAGATTGGTACATTTTACATTTTACATTTTACATTTTACATTTTACATTTTACATTTT 1389
QY 533 TAGACGACCGCTACTTACTTACTTCCGCGAGAAACCTTCTGGAAGACGCGCTCAAAAGGG 592
Db 1388 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1329
```

```
QY 593 AAGATACGAGTATGTTTGGAGCACACTTTCAAAATCGGTATGCGAGTTTAAGGGTAGGAA 652
Db 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1269
QY 653 TTTTATAGGGAAGGGGAAATCGTCTCAGAAATTAACAAGGTGAGACTCCCTTCAACG 712
Db 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1209
QY 713 TGACCTACCCCTCTCAGGTGATGCAAAAGTTCTCTCAGGAGGGAAGAGAAATTCCTTAA 772
Db 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1149
QY 773 TCGAAAGATACAGAGGTTGTACAGAGCGGAGAAAGATGTACAGCAAAATGAAGAAA 832
Db 1148 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1089
QY 833 TAGAAGGAGTGTAGG 847
Db 1088 RRRRRRRRRRRR 1074

RESULT 6
US-08-646-590B-35
; Sequence 35, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1092 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1089
US-08-646-590B-35
```









QY 387 TGATATAGACTTAGAAGAAAGTATTGAATT 416  
 Db 211 TGCATAAATTCAAAAATATTGTGGGTTT 182

## RESULT 13

US-08-558-865-7/c  
 ; Sequence 7, Application US/08558865  
 ; Patent No. 5750386  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Conkling, Mark A.  
 ; APPLICANT: Opperman, Charles H.  
 ; APPLICANT: Taylor, Christopher G.  
 ; TITLE OF INVENTION: Pathogen-Resistant Transgenic Plants  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and  
 ; STREET: Post Office Drawer 34009  
 ; CITY: Charlotte  
 ; STATE: No. 5750386th Carolina  
 ; COUNTRY: U.S.A.  
 ; ZIP: 28234

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/558,865  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/236,678  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; APPLICATION NUMBER: US/07/770,082  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sibley, Kenneth D.  
 ; REGISTRATION NUMBER: 31,665  
 ; REFERENCE/DOCKET NUMBER: 5051-166  
 ; TELEPHONE: 919-881-3140  
 ; TELEFAX: 919-881-3175  
 ; TELEX: 575102

INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1100 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 US-08-558-865-7

Query Match 3.1%; Score 33.2; DB 1; Length 1100;  
 Best Local Similarity 51.3%; Pred. No. 0.88;  
 Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 267 AATAGCTATAGTGAACCTTACATACCCGTTTACATACCTGTTCACACCTTTCCCATGTA 326  
 Db 331 ATTTGATTAAGATGCTCCACATCCCTTTTACAAACTATTCTTACTTTTGGTTC 272  
 QY 327 CGAGATAAGTCGGAAGCTTCTCGGAAGACCCCTCGTAAGGTTCACTCGAGCAAACTT 386  
 Db 271 TTAGATTGTTGTAATTTCTTAAGGAAATAGTTGCATCAAGCGCATTTCTAAAGTA 212  
 QY 387 TGATATAGACTTAGAAGAAAGTATTGAATT 416  
 Db 211 TGCATAAATTCAAAAATATTGTGGGTTT 182

## RESULT 14

US-08-234-939-6/c  
 ; Sequence 6, Application US/08234939  
 ; Patent No. 5459252  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Conkling, Mark A.  
 ; APPLICANT: Yamamoto, Yuri T.  
 ; TITLE OF INVENTION: Root Specific Gene Promoter  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson  
 ; STREET: Post Office Drawer 34009  
 ; CITY: Charlotte  
 ; STATE: No. 5459252th Carolina  
 ; COUNTRY: U.S.A.  
 ; ZIP: 28234

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.24  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/234,939  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/649,564  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sibley, Kenneth D.  
 ; REGISTRATION NUMBER: 31,665  
 ; REFERENCE/DOCKET NUMBER: 5051-141  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 919-881-3140  
 ; TELEFAX: 919-881-3175  
 ; TELEX: 575102

INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1268 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 US-08-234-939-6

Query Match 3.1%; Score 33.2; DB 1; Length 1268;  
 Best Local Similarity 51.3%; Pred. No. 0.96;  
 Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 267 AATAGCTATAGTGAACCTTACATACCCGTTTACATACCTGTTCACACCTTTCCCATGTA 326  
 Db 499 ATTTGATTAAGATGCTCCACATCCCTTTTACAAACTATTCTTACTTTTGGTTC 440  
 QY 327 CGAGATAAGTCGGAAGCTTCTCGGAAGACCCCTCGTAAGGTTCAACTGGACGAAACTT 386  
 Db 439 TTAGATTGTTGTAATTTCTTAAGGAAATAGTTGCATCAAGCGCATTTCTAAAGTA 380  
 QY 387 TGATATAGACTTAGAAGAAAGTATTGAATT 416  
 Db 379 TGCATAAATTCAAAAATATTGTGGGTTT 350

## RESULT 15

US-08-558-865-6/c  
 ; Sequence 6, Application US/08558865  
 ; Patent No. 5750386  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Conkling, Mark A.  
 ; APPLICANT: Opperman, Charles H.  
 ; APPLICANT: Taylor, Christopher G.  
 ; TITLE OF INVENTION: Pathogen-Resistant Transgenic Plants  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and

us-09-905-173-23.rni

Wed Apr 16 08:05:51 2003

ADDRESSEE: Gibson  
 STREET: Post Office Drawer 34009  
 CITY: Charlotte  
 STATE: No. 5750386th Carolina  
 COUNTRY: U.S.A.  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/558,865  
 FILING DATE:  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/236,678  
 FILING DATE:  
 CLASSIFICATION: 800  
 APPLICATION NUMBER: US/07/770,082  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 REFERENCE/DOCKET NUMBER: 5051-166  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-881-3140  
 TELEFAX: 919-881-3175  
 TELEX: 575102  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1268 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-558-865-6

Query Match 3.1%; Score 33.2; DB 1; Length 1268;  
 Best Local Similarity 51.3%; Pred. No. 0.96;  
 Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 Qy 267 AATAGCTATAGTGAACCTTTACATACCGTTTACATACCTGTTCCACCTTTCCCATGTA 326  
 Db 499 ATTTGATTAGATGCTCCTCCACATCCCTTTTACAACTATTTTCTACTTTTGTGATGTC 440  
 Qy 327 CGAGATAGTCCGAAGCTTCTCGAAGACCCCTCGTAAAGGTTCAACTGGACGAGAAACTT 386  
 Db 439 TTAGATTGTTGTGAATATTTCCCTAAGGAAATAGTTGCCATCAACCCATTTCTAAAGTA 380  
 Qy 387 TGATATAGACTTAGAAGAAAGTATTGAATT 416  
 Db 379 TGCCATAAATTCAAAAATATTTGTGGGTTT 350

Search completed: April 15, 2003, 15:49:10  
 Job time : 100 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 15:15:16 : Search time 125 seconds  
(without alignments)  
7473.469 Million cell updates/sec

Title: US-09-905-173-23  
Perfect score: 1065  
Sequence: 1 atgataccacagaggattaa.....ccctttcaagctctcttttaa 1065

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438593890 residues  
Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/PCTUS\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065	100.0	1065	9 US-10-060-432-23	Sequence 23, Appl
2	1065	100.0	1065	10 US-09-905-173-23	Sequence 23, Appl
3	46.2	4.3	1085	10 US-09-974-300-298	Sequence 298, Appl
4	42.2	4.0	1092	10 US-09-905-173-35	Sequence 35, Appl
5	37.8	3.5	198	9 US-10-083-357-579	Sequence 579, Appl
6	37.6	3.5	588	10 US-09-864-761-6646	Sequence 6646, Ap
7	36.2	3.4	6804	7 US-08-781-986A-365	Sequence 365, App
8	35.8	3.4	560	10 US-09-974-300-4716	Sequence 4716, Ap
9	35.8	3.4	1098	9 US-09-738-626-2298	Sequence 2298, Ap
10	35.2	3.3	1350	9 US-09-971-536-8	Sequence 8, Appli
11	34.6	3.2	2000	9 US-09-938-842A-3936	Sequence 3936, Ap
12	34.4	3.2	375	10 US-09-864-761-11114	Sequence 11114, A
13	34.4	3.2	2270	10 US-09-887-576-455	Sequence 455, App
14	34.4	3.2	12225	10 US-09-927-668-11	Sequence 11, Appl
15	34.4	3.2	12616	10 US-09-927-668-9	Sequence 9, Appli
16	34.2	3.2	534	9 US-09-738-626-2011	Sequence 2011, Ap
17	34.2	3.2	1093	9 US-10-001-876-78	Sequence 78, Appl
18	34.2	3.2	1691139	9 US-10-067-514-1	Sequence 1, Appli
19	34.2	3.2	3309400	9 US-09-738-626-1	Sequence 1, Appli

20	33.8	3.2	671	9 US-10-184-644-346	Sequence 346, App
21	33.6	3.2	1851	10 US-09-974-300-2020	Sequence 2020, Ap
22	33.2	3.1	1023	9 US-09-738-626-245	Sequence 245, App
23	33.2	3.1	1480	10 US-09-948-649-1	Sequence 1, Appli
24	33	3.1	837	10 US-09-974-300-7726	Sequence 7726, Ap
25	33	3.1	948	10 US-09-974-300-7714	Sequence 7714, Ap
26	32.8	3.1	1757	9 US-09-938-842A-4699	Sequence 4699, Ap
27	32.6	3.1	319	10 US-09-974-300-1427	Sequence 1427, Ap
28	32.6	3.1	446	10 US-09-864-761-1815	Sequence 1815, Ap
29	32.6	3.1	53332	9 US-10-224-562-3	Sequence 3, Appli
30	32.6	3.1	53332	10 US-09-801-861-3	Sequence 3, Appli
31	32.4	3.0	668	7 US-08-781-986A-309	Sequence 309, App
32	32.4	3.0	1872	10 US-09-815-242-4336	Sequence 4336, Ap
33	32.4	3.0	1884	10 US-09-815-242-8207	Sequence 8207, Ap
34	32.2	3.0	477	12 US-10-028-780-12	Sequence 12, Appl
35	32	3.0	336	10 US-09-969-708-449	Sequence 449, App
36	32	3.0	336	10 US-09-969-347-312	Sequence 312, App
37	31.8	3.0	472	9 US-10-184-644-98	Sequence 98, Appl
38	31.6	3.0	161	10 US-09-864-761-23380	Sequence 23380, A
39	31.6	3.0	5134	10 US-09-867-852-157	Sequence 157, App
40	31.6	3.0	15772	10 US-09-764-903-66	Sequence 66, Appl
41	31.6	3.0	32768	10 US-09-070-927A-100	Sequence 100, App
42	31.6	3.0	32816	10 US-09-729-094-3	Sequence 3, Appli
43	31.4	2.9	242	10 US-09-864-761-21849	Sequence 21849, A
44	31.4	2.9	392	9 US-10-184-644-160	Sequence 160, App
45	31.4	2.9	498	10 US-09-864-761-5128	Sequence 5128, Ap

## ALIGNMENTS

RESULT 1  
US-10-060-432-23  
: Sequence 23, Application US/10060432  
: Publication No. US20030040092A1  
: GENERAL INFORMATION:  
: APPLICANT: WARREN, Patrick V.  
: SWANSON, Ronald V.  
: TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
: NUMBER OF SEQUENCES: 32  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: CARELLA, STEWART & OLSTEIN  
: CECCHI, STEWART & OLSTEIN  
: STREET: 6 BECKER FARM ROAD  
: CITY: ROSELAND  
: STATE: NEW JERSEY  
: COUNTRY: USA  
: ZIP: 07068  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5 INCH DISKETTE  
: OPERATING SYSTEM: MS-DOS  
: SOFTWARE: WORD PERFECT 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/10/060.432  
: FILING DATE: 29-Jan-2002  
: CLASSIFICATION: <Unknown>  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 09/481,733  
: FILING DATE: 11-JAN-2000  
: APPLICATION NUMBER: US/08/599,171  
: FILING DATE: <Unknown>  
: ATTORNEY/AGENT INFORMATION:  
: NAME: HERRON, CHARLES J.  
: REGISTRATION NUMBER: 28,019  
: REFERENCE/DOCKET NUMBER: 331400-38  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 201-994-1700  
: TELEFAX: 201-994-1744  
: INFORMATION FOR SEQ ID NO: 23:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1065 NUCLEOTIDES  
: TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
US-10-060-432-23

Query Match 100.0%; Score 1065; DB 9; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60  
DB 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60

QY 61 GTCAGGCTTCTCTACCAAAATCCCTACGACTTCCCGAGAGATAAACAAAGGCC 120  
DB 61 GTCAGGCTTCTCTACCAAAATCCCTACGACTTCCCGAGAGATAAACAAAGGCC 120

QY 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAAATACCCAGACCCGGAAGAGAGTTA 180  
DB 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAAATACCCAGACCCGGAAGAGAGTTA 180

QY 181 AAAGCGGTTCTGCGGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCGGTAACGGT 240  
DB 181 AAAGCGGTTCTGCGGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCGGTAACGGT 240

QY 241 TCGGAGAACTCATATACCTCTCAATAGCTATAGTGAACTTACATACCGGTTTAC 300  
DB 241 TCGGAGAACTCATATACCTCTCAATAGCTATAGTGAACTTACATACCGGTTTAC 300

QY 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360

QY 361 GTAAAGGTTCAACTGCAGCAAACTTTGATATAGCTTGAAGAAGTATTCGAATTAATA 420  
DB 361 GTAAAGGTTCAACTGCAGCAAACTTTGATATAGCTTGAAGAAGTATTCGAATTAATA 420

QY 421 GAGAAAGAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAACCTTC 480  
DB 421 GAGAAAGAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAACCTTC 480

QY 481 TTTTCAGGGGAAGATGAGGAGATAAGAACAGGGGTGTTCTGTCTGTAATAGACGAA 540  
DB 481 TTTTCAGGGGAAGATGAGGAGATAAGAACAGGGGTGTTCTGTCTGTAATAGACGAA 540

QY 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGAAGATAC 600  
DB 541 GCCTACTATCATTTACTCCGAGAAACCTTTCTGGAAGACGCGCTCAAAAGGAAGATAC 600

QY 601 GTAGTTTGAAGACATTTTCAAAATCGGTATGCGAGTTTAAAGGTAGGATTTTAATA 660  
DB 601 GTAGTTTGAAGACATTTTCAAAATCGGTATGCGAGTTTAAAGGTAGGATTTTAATA 660

QY 661 GGGAGGGGAATTCGCTCAGAAATTAACAAGTGAGACTCCCTTCAAGGTGACCTAC 720  
DB 661 GGGAGGGGAATTCGCTCAGAAATTAACAAGTGAGACTCCCTTCAAGGTGACCTAC 720

QY 721 CCCTCTCAGGTGATGGCAAAAGTTCTCCTCACGGAGGAGAGAAATTCCTAATGGAAG 780  
DB 721 CCCTCTCAGGTGATGGCAAAAGTTCTCCTCACGGAGGAGAGAAATTCCTAATGGAAG 780

QY 781 ATACAGAGGTTGTAACAGAGCGAGAAAGGATGTACACGAAATGAAGAAATAGAAGA 840  
DB 781 ATACAGAGGTTGTAACAGAGCGAGAAAGGATGTACACGAAATGAAGAAATAGAAGA 840

QY 841 GTTCAGGTTTTCGAGTAAAGGCTTAACCTTCTGCTTTTCAAGACGCTTACCCGCCAC 900  
DB 841 GTTCAGGTTTTCGAGTAAAGGCTTAACCTTCTGCTTTTCAAGACGCTTACCCGCCAC 900

QY 901 GAGGTTTATCAGGAGCTACTGAAAAGGATGTCTCTCGTCAGGAACGATCTTACATGGAA 960  
DB 901 GAGGTTTATCAGGAGCTACTGAAAAGGATGTCTCTCGTCAGGAACGATCTTACATGGAA 960

QY 961 GGACTCCAAAAGTCCCTCAGGTAAGCGTAGGGAACCGGAAGAAACAAACAAAGTTTCTG 1020  
DB 961 GGACTCCAAAAGTCCCTCAGGTAAGCGTAGGGAACCGGAAGAAACAAACAAAGTTTCTG 1020

QY 1021 GAACACTGAGGAGAGATATAAATCCCTTCAAGCTCTCTTTAA 1065  
DB 1021 GAACACTGAGGAGAGATATAAATCCCTTCAAGCTCTCTTTAA 1065

RESULT 2  
US-09-905-173-23  
; Sequence 23, Application US/0905173  
; Patent No. US20020132295A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay M.  
; APPLICANT: WARREN, Patrick V.  
; APPLICANT: SWANSON, Ronald V.  
; APPLICANT: MATHUR, Eric J.  
; TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND  
; FILE REFERENCE: DIVER1240-7  
; CURRENT APPLICATION NUMBER: US/09/905,173  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: US 09/412,184  
; PRIOR FILING DATE: 1999-10-04  
; PRIOR APPLICATION NUMBER: US 09/389,537  
; PRIOR FILING DATE: 1999-09-02  
; PRIOR APPLICATION NUMBER: US 08/646,590  
; PRIOR FILING DATE: 1996-05-08  
; PRIOR APPLICATION NUMBER: US 08/599,171  
; PRIOR FILING DATE: 1996-02-09  
; PRIOR APPLICATION NUMBER: US 09/481,733  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: US 09/069,226  
; PRIOR FILING DATE: 1998-04-27  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1065  
; TYPE: DNA  
; ORGANISM: Aquifex  
US-09-905-173-23

Query Match 100.0%; Score 1065; DB 10; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60  
DB 1 ATGATACCCAGAGGATTAAAGAACTTGAAGCTTACAAGACGAGGTCACTCCGCCCTCC 60

QY 61 GTCAGGCTTCTCTACCAAAATCCCTACGACTTCCCGAGAGATAAACAAAGGCC 120  
DB 61 GTCAGGCTTCTCTACCAAAATCCCTACGACTTCCCGAGAGATAAACAAAGGCC 120

QY 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAAATACCCAGACCCGGAAGAGAGTTA 180  
DB 121 TTAGAAGAATTAATAAAGGTTCCCTTGAACAAATACCCAGACCCGGAAGAGAGTTA 180

QY 181 AAAGCGGTTCTGCGGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCGGTAACGGT 240  
DB 181 AAAGCGGTTCTGCGGATTTTTCGGCGTTAAGGAAGAAATTTAGTTCGGTAACGGT 240

QY 241 TCGGAGAACTCATATACCTCTCAATAGCTATAGTGAACTTACATACCGGTTTAC 300  
DB 241 TCGGAGAACTCATATACCTCTCAATAGCTATAGTGAACTTACATACCGGTTTAC 300

QY 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360  
DB 301 ATACCTGTTCCCACTTCCCATGTACGAGATAAGTGCAGAAAGTTCTCGGAAGACCCCTC 360

QY	361	GTAAAGGTTCACTGCACGAAACCTTTGATATAGACTTTAGAAAGAAGTATTGAATTAATA	420
Db	361	GTAAAGGTTCACTGCACGAAACCTTTGATATAGACTTTAGAAAGAAGTATTGAATTAATA	420
QY	421	GAGAAAGAAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAAACCTC	480
Db	421	GAGAAAGAAAACCCGTTCTCGGGTACTTTGCTTACCCAAACACCCACGGGAAACCTC	480
QY	481	TTTTTCAGGGGAAAGATTGAGGAGATAGAAACACAGGGGTGTTTCTGTGTAATAGACGAA	540
Db	481	TTTTTCAGGGGAAAGATTGAGGAGATAGAAACACAGGGGTGTTTCTGTGTAATAGACGAA	540
QY	541	GCCTACTATCATTTACTTCGAGAGAAACCTTCTCGAAGACGCGCTCAAAAGGGAAGATACG	600
Db	541	GCCTACTATCATTTACTTCGAGAGAAACCTTCTCGAAGACGCGCTCAAAAGGGAAGATACG	600
QY	601	GTAGTTTTGAGGACACTTTCAAAATCGGTATGCGGAGTTTAAAGGTAGGATTTTAATA	660
Db	601	GTAGTTTTGAGGACACTTTCAAAATCGGTATGCGGAGTTTAAAGGTAGGATTTTAATA	660
QY	661	GGAAAGGGGAAATCGTCTCAGAAATTACAAAGGTGAGACTCCCTTCAACGTCACCTAC	720
Db	661	GGAAAGGGGAAATCGTCTCAGAAATTACAAAGGTGAGACTCCCTTCAACGTCACCTAC	720
QY	721	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAGGAGGGAAGAGAAATTCCTAATGGAAAG	780
Db	721	CCCTCTCAGGTGATGGCAAAAGTTCTCTCAGGAGGGAAGAGAAATTCCTAATGGAAAG	780
QY	781	ATACAGGAGGTCTACACAGCGAGAAAGGATGTACGACGAAATGAGAAATAGAGGA	840
Db	781	ATACAGGAGGTGTAAACAGAGCGAGAAAGGATGTACGACGAAATGAGAAATAGAGGA	840
QY	841	GTTGAGGTTTTCCGAGTAGAGCTAACTCTTCTGCTTTTCAGAACGCCCTTACCCGCCAC	900
Db	841	GTTGAGGTTTTCCGAGTAGAGCTAACTCTTCTGCTTTTCAGAACGCCCTTACCCGCCAC	900
QY	901	GAGGTTTATCAGGAGCTACTGAAAAGGATGTCCGTCGTCAGAACGTATCTTACATGAA	960
Db	901	GAGGTTTATCAGGAGCTACTGAAAAGGATGTCCGTCGTCAGAACGTATCTTACATGAA	960
QY	961	GGACTCCAAAAGTGCTCAGGCTAAGCGGTAGGGAACCGGAAGAAACACCAAGTTTCTG	1020
Db	961	GGACTCCAAAAGTGCTCAGGCTAAGCGGTAGGGAACCGGAAGAAACACCAAGTTTCTG	1020
QY	1021	GAAGCACTGGAGAGATATAAATCCCTTTCAAGCTCTCTTTAA	1065
Db	1021	GAAGCACTGGAGAGATATAAATCCCTTTCAAGCTCTCTTTAA	1065

## RESULT 3

US-09-974-300-298  
; Sequence 298, Application US/09974300  
; Patent No. US20020146721A1

```

/ PATENT NO.: US20020146721A1
/ GENERAL INFORMATION:
/ APPLICANT: Berka, Randy M.
/ APPLICANT: Clausen, Ib Groth
/ TITLE OF INVENTION: Methods For Monitoring Multiple Gene
/ TITLE OF INVENTION: Expression
/ FILE REFERENCE: 10085.500-US
/ CURRENT APPLICATION NUMBER: US/09/974,300
/ CURRENT FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 09/680,598
/ PRIOR FILING DATE: 2000-10-06
/ PRIOR APPLICATION NUMBER: 60/279,526
/ PRIOR FILING DATE: 2001-03-27
/ NUMBER OF SEQ ID NOS: 8481
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 298
/ LENGTH: 1085
/ TYPE: DNA
/ ORGANISM: Bacillus licheniformis
/ US-09-974-300-298

```

	Query Match	4.3%;	Score 46.2;	DB 10;	Length 1085;
	Best Local Similarity	47.4%;	Pred. No. 0.00077;		
	Matches 138;	Conservative 0;	Mismatches 153;	Indels 0;	Gaps 0;
Qy	154	TACCACGACCCGGAAGCGAAGAGCTTAAAGCGGTTCTTGGGATTTTTCGGCGTTAAG	213		
Db	181	TATCCTGACGCTATAGTGCGCATTGAGAACAAAGCTTGCCTCCCATCTCGGCGTAAT	240		
Qy	214	GAAGAAATTTAGTCTTCGGTAAAGGTTTCGGACGAACCTCATATPACTACCTCTCAATAGCT	273		
Db	241	GAACAGCAATCATCTCTGGGAACCGAACGACGAAGTCATPCTAAATATTTCCTCCGTCT	300		
Qy	274	ATAGGTGAACPTTACATACCCGTTTACATACCTCTTCCCACCTTTCCTCATGTACGAGATA	333		
Db	301	TTGTTGACCTCGCTCAACACACTGTATGCGAATCCGACATTTTCACAATACAAGCAC	360		
Qy	334	AGTCGAAAGTTCFCGGAAGACCCCTCGTAAAGGTTTCAACTGGACGAAACTTTGATATA	393		
Db	361	AACCGTGTATCGAAGGAGCGGAAGTCAGAGAGTCGGCGTCTCGAAACGGCTGCCAC	420		
Qy	394	GACTTAGAAGAAGATTGTAATTTAATAGAGAAAGAAAACCCGTTCTCGGG	444		
Db	421	GACCTTGATGCAATGCTGGAAGCCGATCGATGAACAGACGAAGGTGTGCTGG	471		

## RESULT 4

US-09-905-173-35  
; Sequence 35, Application US/09905173  
; Patent No. US20020132295A1

```

: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: SHORT, Jay M.
: APPLICANT: WARREN, Patrick V.
: APPLICANT: SWANSON, Ronald V.
: APPLICANT: MATHUR, Eric J.
: TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND
: TITLE OF INVENTION: USE THEREOF
: FILE REFERENCE: DIVER124D-7
: CURRENT APPLICATION NUMBER: US/09/905,173
: CURRENT FILING DATE: 2001-07-12
: PRIOR APPLICATION NUMBER: US 09/412,184
: PRIOR FILING DATE: 1999-10-04
: PRIOR APPLICATION NUMBER: US 09/389,537
: PRIOR FILING DATE: 1999-09-02
: PRIOR APPLICATION NUMBER: US 08/646,590
: PRIOR FILING DATE: 1996-05-08
: PRIOR APPLICATION NUMBER: US 08/599,171
: PRIOR FILING DATE: 1996-02-09
: PRIOR APPLICATION NUMBER: US 09/481,733
: PRIOR FILING DATE: 2000-01-11
: PRIOR APPLICATION NUMBER: US 09/069,226
: PRIOR FILING DATE: 1998-04-27
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 35
: LENGTH: 1092
: TYPE: DNA
: ORGANISM: Ammonifex degensii
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (986)..(987)
: OTHER INFORMATION: n is any nucleotide
US-09-905-173-35

```

	Query Match	4.0%	Score 42.2;	DB 10;	Length 1092;
	Best Local Similarity	48.2%;	Pred. No. 0.013;		
	Matches 119;	Conservative	0;	Mismatches 128;	Indels 0;
	Gaps	0;			
QY	138	GGTTCCCTTGAACAAATACCCAGAGACCCCGAAGCGAAAGAGGTAAACGCGGTCTTTCGGGA	197		
Db	168	GGACCATGGACGCTTTACCCAGGAAGGCTCAAGCTATGAGCTACGGCAGCGCTGGGTAA	227		
QY	198	TTTTTTCGGGGTTAAGGAAGAAATTTAGTTCTCGGTAAAGGGTTTCGACGCAACATCATATA	257		

100



```
QY 415 TTAATAGAGAAAGAAAACCCGTTCTCGGTACTTTGCTTACCCAAACACCCACGGGA 474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 GAAATTCGTGCAAGACGCTGACATTTGTTTGTACACACCCCGAACAACCCGACCGT 525
QY 475 AACCTCTTTTCAGGGGAAGATTGAGGAGATAAGAAAC 513
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 GATGTGACCTCGCTGGACGATGTTGAGCCATCAACAAC 564

RESULT 10
US-09-971-536-8
; Sequence 8, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-8

Query Match 3.3%; Score 35.2; DB 9; Length 1350;
Best Local Similarity 46.1%; Pred. No. 2.1;
Matches 155; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 152 AATACCAGACCCCGAAGCGGTAACGAGTTAAACCGGTTCTTCGGGATTTTTCGGGGTTA 211
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 ATTACCCCGATGCTATGCCAGTCAACTACGCCCGGTTGGCAAAACATTTAGACGTTG 363
QY 212 AGGAAGAAAATTTAGTTCTCGGTAACGTTTCGGACGAACCTCATATACTACCTCTCAATAG 271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 CCGAGAGACGTTGGTTTTGGCAATGGTCTGATGAATCATTCCTTAATTCGCCGCA 423
QY 272 CTATAGTGAACCTTACATACCCGTTTACATACCTGTTCACCTTCCCATGTACGAGA 331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 CTTTTTTGAGCCCGGGGGATGAAGTCATTCACCATGGCCAAACATTTCCGAGTACCGCT 483
QY 332 TAAGTGGAAAGTTCTCGGAAGACCCCTCGTAAGGTTCACTGGAGAAACCTTTGATA 391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 TGCATGCCCAATTTGAAGGGGCCACCGTGATGATGTGCCCGTCACCTGAACACTGGCAAT 543
QY 392 TAGACTTAGAAGAAAGTATTGAATTAATACAGAAAGAAAACCCGTTCTCGGGTACTTTG 451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 TTGATTTATCTGCANTGGCCAGCGGCTA---ACCGGAAAACGAACGATTTGGGTGT 600
QY 452 CTTACCAAAACACCCCGACGGGAACCTCTTTTCCA 487
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 GCAACCAAAATAACCCCGACGGCGCTGCTGTCAA 636

RESULT 11
```

```
US-09-938-842A-3936/c
; Sequence 3936, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3936
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3936

Query Match 3.2%; Score 34.6; DB 9; Length 2000;
Best Local Similarity 59.8%; Pred. No. 4.1;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 822 AATGAAGAAATAGAGAGTTGAGGTTTTCGAGTAAGCTAAGCTTCTGCTTTTTCAG 881
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1625 AAAACAGAAATTGAGAAGGTTTGGGAAGATCTCCAAATAAAGTCAACGCTGCTATTTTTCAC 1566
QY 882 AAGCGCTTACCCCGCCAGGAGTTTATCAGGAGCTA 918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1565 AAGTTTAAATCTGGCTAATAGGTTTATGATCATCTA 1529

RESULT 12
US-09-864-761-11114
; Sequence 11114, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```





	Matches	98:	Conservative	0:	Mismatches	106:	Indels	0:	Gaps	0:
Qy	102	GGAGATAAAACAAAGGCCTTAGAAGAAATTA	AAAAAAAAAGTTTCCCTTGAAACAAATACCCAGA	161						
Db	8817	GGATTCAAAATCA	AAAGGATATATCTAA	AATAGTGCAGATATCACCCAGGCAGTGTCTCT	8876					
Qy	162	CCCCGAAGCGAAAGAGTTTAAAGCGGTTCTTGC	CGGATTTTTCGCGCTTAGGGAAGAAA	221						
Db	8877	CTCCCAAGAAATCAGAGAAAAGGTGATCCAGCAT	TATTAGAGGAATGTATAAAGTAGA	8936						
Qy	222	TTTAGTTCCTCGGTAAACGGTTCGGAGCACTCAT	ATATACTACCTCTCAATAGCTATAGGTGA	281						
Db	8937	TTTGAGTGCCAGCAGACATTGCAGGAAC	TATTTCAGCAGCTGCACATGATAGAGCAGT	8996						
Qy	282	ACTTTACATACCCGTTTACATACC	305							
Db	8997	ATGGTATGACCCCATCTACTATCC	9020							

Search completed: April 15, 2003, 16:58:26  
Job time : 144 secs

[illegible]

```

RESULT 15
US-09-927-668-9
; Sequence 9, Application US/09927668
; Patent No. US20020115144A1
; GENERAL INFORMATION:
; APPLICANT: Kaplan, Jerry
; Mourou, Karen
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/927,668
; FILING DATE: 10-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/396,540
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-062-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12616 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 190..11592
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-927-668-9

```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 13:59:36 ; Search time 2156 seconds

(without alignments)

8000.090 Million cell updates/sec

Title: US-09-905-173-23

Perfect score: 1065

Sequence: 1 atgatacccccaggaggattaa.....ccctttcaagctctcttttaa 1065

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pin:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.8	6.2	573	10	BE432349 EST398878
2	65	6.1	587	10	AW034522 EST278138
3	62.6	5.9	819	12	BG350830 098611 Ma
4	58.8	5.5	762	13	BM412207 EST386534
5	58.4	5.5	631	13	BJ463860 BJ463860
6	57.2	5.4	778	12	BG887294 EST513145

7	53.4	5.0	811	12	BG598806
8	52.4	5.0	622	13	BM322300
9	52.4	4.9	639	14	BM812973
10	51.6	4.8	555	13	BJ280409
11	50.2	4.7	405	13	BM173372
12	47.8	4.5	511	13	BI1717525
13	47.2	4.4	667	14	B0115402
14	47.2	4.4	1757	11	AV105176
15	46.2	4.3	662	13	BM003279
16	45.6	4.3	997	17	CNS005TE
17	44.6	4.2	940	17	CNS06YZU
18	44.2	4.2	925	17	CNS00IBN
19	43.6	4.1	299	9	AI423192
20	43	4.0	313	10	BE341733
21	42.6	4.0	515	13	BI934083
22	42.6	4.0	659	13	BI934806
23	41.6	3.9	540	10	AW065601
24	41.6	3.9	1101	17	CNS017KE
25	40.6	3.8	322	9	AI276240
26	40.2	3.8	588	13	BI994014
27	40	3.8	532	13	BJ128028
28	40	3.8	556	13	BJ108720
29	40	3.8	625	13	BJ104535
30	40	3.8	658	13	BJ120960
31	39.8	3.7	1101	17	CNS0106X
32	39.4	3.7	860	17	A2678707
33	39.4	3.7	889	17	A2677419
34	39.4	3.7	901	17	BI161744
35	39.4	3.7	902	17	A2543474
36	39.4	3.7	920	17	BI151132
37	39.2	3.7	418	12	BG509907
38	39.2	3.7	565	14	B0134497
39	38.4	3.6	612	13	BJ117344
40	38.4	3.6	631	13	BJ114619
41	38.4	3.6	1101	17	CNS00G3K
42	37.8	3.5	583	14	BQ123162
43	37.8	3.5	873	17	CNS018MF
44	37.8	3.5	898	17	BI133728
45	37.6	3.5	1101	17	CNS00BJ6

## ALIGNMENTS

RESULT 1	BE432349	EST398878	tomato breaker fruit, TIGR Lycopersicon esculentum cDNA	573 bp	linear	EST 18-MAY-2001
LOCUS	BE432349	clone cLEG7J1, mRNA sequence.				
DEFINITION	BE432349	EST				
ACCESSION	BE432349	GI:9430192				
VERSION	EST					
KEYWORDS	tomato.					
SOURCE	Lycopersicon esculentum					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;					
REFERENCE	1 (bases 1 to 573)					
AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.					
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: CUGI					
	Clemson University Genomics Institute					
	Clemson University					
	100 Jordan Hall, Clemson, SC 29634, USA					
	Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a>					
FEATURES	5 prime sequence.					
	Location/Qualifiers					
	1..573					



# **AUTHORS** **TITLE** **JOURNAL** **COMMENT**

Nielsen, K.L., Crookshanks, M., Emmersen, J. and Welinder, K.G.  
EST-sequencing of mature potato tuber (Var. Kuras)  
Unpublished (2000)  
Contact: Karen G. Welinder  
Institut for bioteknologi  
Aalborg Universitet  
Sohngaardsholmsvej 49, 9000 Aalborg, Denmark  
Tel: +45 96358467  
Fax: +45 98141808  
Email: kw@bio.auc.dk  
Sequenced from the 5' end.  
High quality sequence stop: 819  
POLYA-No.

## **FEATURES** **source**

Location/Qualifiers  
1. .819  
/organism="Solanum tuberosum"  
/cultivar="Field grown Kuras"  
/db\_xref="taxon:4113"  
/clone\_lib="Mature tuber lambda ZAP"  
/tissue\_type="Tuber"  
/note="Vector: Lambda ZAP"  
227 a 155 c 194 g 243 t

### **Query Match**

Best Local Similarity 5.9%; Score 62.6; DB 12; Length 819;  
Matches 194; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 149 ACAATACCCAGACCCGACGAAAGAGTTAAAGCGGTCTTCGCGGATTTTCGGCG 208  
DB 393 ATATTATCTCGATCTCGTGAAGACCGGTGCGTGCAGCTCTTGCTGAAGATTCCTGCC 452  
QY 209 TTAAGGAAGAAATTTAGTTCTCGTAAACGTTCCGGACGAACTCATATACCTCTCAA 268  
DB 453 TTGAATCTGAGTATATCTTCAGGTCGCGTGCGAGTGAACCTCATTTGATATGA 512  
QY 269 TAGCTATAGGTGAACCTTTACATACCGGTTTACATACCTGTTCCACCTTTCCCATGACG 328  
DB 513 GATCGTATGGATCTCGTGTGACAAGATTTGTTGACTGCCACCCACTTTCACAATGATG 572  
QY 329 AGATAAGTGGCAAGTTCTCGGAAGACCCCTCGTAAAGGTTCACTGACGACAAACTTTG 388  
DB 573 AATTGATGACGCTGTAATGGACAGGTTGTCATCAAGGTGCTAGGAACCCAGACTTTA 632  
QY 389 ATATAGACTTGAAGAAGTATTGAATTAATAGAGAAGAAACCCGTTCTCGGGTACT 448  
DB 633 GCCTGGATGTAGACGATTCGCCAGTGGTGAACCTTGACAAACCGAAGTGCATATTC 692  
QY 449 TTGCTTACCCAAACACCCGGAACCTCTTTTCCAGGGGAAAGATTGAGGAGATAA 508  
DB 693 TGACATCACCTAATATCTCTGATGGAGTATTAATTGATGATGAACCTTATTGAAAATAC 752  
QY 509 GAAACAGGGGTGTTTCTGTGTTATAGACGACGCTACTATCATCTACTCCGGA 561  
DB 753 TTCACCTGCCCATATTGGTAAATATGGATGAAGCATATGTTGAGTTTCTCGA 805

### **RESULT 4**

BM412207 762 bp mRNA linear EST 22-JAN-2002  
LOCUS EST586534 tomato breaker fruit Lycopersicon esculentum cDNA clone  
DEFINITION CLEGS9E12 5' end, mRNA sequence.  
ACCESSION BM412207  
VERSION BM412207.1 GI:18263837  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 762)

## **REFERENCE** **AUTHORS**

Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai

# **TITLE** **JOURNAL** **COMMENT**

J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Ronning  
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.

## **FEATURES** **source**

Location/Qualifiers  
1. .762  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEGS9E12"  
/clone\_lib="tomato breaker fruit"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSMCUadapt; Site\_1: EcoRI;  
Site\_2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."  
207 a 155 c 178 g 222 t

## **BASE COUNT** **ORIGIN**

Query Match 5.5%; Score 58.8; DB 13; Length 762;  
Best Local Similarity 48.8%; Pred. No. 2.9e-05;  
Matches 159; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 149 ACAATACCCAGACCCGACGAAAGAGTTAAAGCGGTCTTCGCGGATTTTTCGGCG 208  
DB 395 ACATTTATCTGTATCTCGAAGCCGCGTCCGTCAGCTCTTGCTGAAGATTCCTGCC 454  
QY 209 TTAAGGAAGAAATTTAGTTCTCGGTAAACGTTTCGGACGAACTCATATACCTCTCAA 268  
DB 455 TTGAATCTGAGTATATCTTTCAGGGTTCGCGTGCAGATGAACCTCATTTGATAATGA 514  
QY 269 TACCTATAGTGAACCTTACATACCCGTTTACATACCTGTTCCACCTTTCCCATGACG 328  
DB 515 GATGCATATTGGATCTCGTGTGACAAGATTGTTGACTGCCACCCACTTTCACAATGATG 574  
QY 329 AGATAAGTGCAGAAAGTTCTCGGAAGACCCCTCGTAAAGGTTCAACTGGACGAAACTTTG 388  
DB 575 AATTTGATGACGCTGTAATGGAGCAGGTGTCATCAAGTGCCTAGGAACCCAGACTTTA 634  
QY 389 ATATAGACTTGAAGAAGTATTGAATTAATAGAGAAGAAACCCGTTTCGGGTACT 448  
DB 635 GCCTGACGCTTGAACGATTTACCGAAGTGTGGAAGCTGAGAAACAAAGTCATATTTCT 694  
QY 449 TTGCTTTACCAACACACCCACGCGGA 474  
DB 695 GACATCACCTAATAATCTGATGGGA 720

## **RESULT 5** **LOCUS**

DEFINITION BJ463860 K. Sato unpublished cDNA library, cv. Haruna Nijo  
germination shoots Hordeum vulgare subsp. vulgare cDNA clone  
bags30m21 5', mRNA sequence.  
ACCESSION BJ463860  
VERSION BJ463860.1 GI:21142367  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare.  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;





Db 352 GTACTTGAACCAAGGTGACAAATTTGATAGTTGCCCTCCAACTTTACCAATGATGAGTTT 411

10

10





Coe, J. C.  
 DIRECT SUBMISSION  
 TITLE  
 Submitted (25-APR-2002) Maize Mapping Project, university of  
 JOURNAL  
 Columbia, MO 65211, USA  
 Missouri,  
 FEATURES  
 Location/Qualifiers  
 1. l757  
 /organism="Zea mays"  
 /db\_xref="MaizeDB:636533"  
 /db\_xref="taxon:4577"  
 /clone="PCO111982"  
 /clone\_lib="Maize Mapping Project/DuPont Consensus



